DTIC® has determined on 10 / JUL / 2013 that this Technical Document has the Distribution Statement checked below. The current distribution for this document can be found in the DTIC® Technical Report Database.

☒ **DISTRIBUTION STATEMENT A.** Approved for public release; distribution is unlimited.

☐ **© COPYRIGHTED.** U.S. Government or Federal Rights License. All other rights and uses except those permitted by copyright law are reserved by the copyright owner.

☐ **DISTRIBUTION STATEMENT B.** Distribution authorized to U.S. Government agencies only (fill in reason) (date of determination). Other requests for this document shall be referred to (insert controlling DoD office).

☐ **DISTRIBUTION STATEMENT C.** Distribution authorized to U.S. Government Agencies and their contractors (fill in reason) (date determination). Other requests for this document shall be referred to (insert controlling DoD office).

☐ **DISTRIBUTION STATEMENT D.** Distribution authorized to the Department of Defense and U.S. DoD contractors only (fill in reason) (date of determination). Other requests shall be referred to (insert controlling DoD office).

☐ **DISTRIBUTION STATEMENT E.** Distribution authorized to DoD Components only (fill in reason) (date of determination). Other requests shall be referred to (insert controlling DoD office).

☐ **DISTRIBUTION STATEMENT F.** Further dissemination only as directed by (insert controlling DoD office) (date of determination) or higher DoD authority.

Distribution Statement F is also used when a document does not contain a distribution statement and no distribution statement can be determined.

☐ **DISTRIBUTION STATEMENT X.** Distribution authorized to U.S. Government Agencies and private individuals or enterprises eligible to obtain export-controlled technical data in accordance with DoDD 5230.25; (date of determination). DoD Controlling Office is (insert controlling DoD office).

AU–18

SPACE HANDBOOK

AIR UNIVERSITY
MAXWELL AIR FORCE BASE, ALABAMA

MATHEMATICAL SIGNS AND SYMBOLS

$\pm$ plus or minus, positive or negative

$\neq$ is not equal to

$\equiv$ is identical to

$\approx$ approximately equal to

$>$ greater than

$\geqq$ greater than or equal to

$<$ less than

$\leqq$ less than or equal to

$\backsim$ similar to

$\propto$ varies as, proportional to

$\rightarrow$ approaches as a limit

∞ infinity

$\therefore$ therefore

$\sqrt{\ }$ square root

$\sqrt[n]{\ }$ nth root

a^n nth power of "a"

a^{-n} reciprocal of nth power of a $= \left(\frac{1}{a^n}\right)$

log, $\log_{10}$ common logarithm

ln, $\log_e$ natural logarithm

n^o n degrees

n' n minutes; n feet

n" n seconds; n inches

$f(x)$ function of x

Δx increment of x

dx differential of x

∂x partial differential of x

Σ summation of

$\int$ symbol for integration

GREEK ALPHABET

Alpha	A	α	Iota	Ι	ι	Rho	Ρ	ρ
Beta	B	β	Kappa	Κ	κ	Sigma	Σ	σ
Gamma	Γ	γ	Lambda	Λ	λ	Tau	Τ	τ
Delta	Δ	δ	Mu	Μ	μ	Upsilon	Υ	υ
Epsilon	Ε	ε	Nu	Ν	ν	Phi	Φ	φ
Zeta	Ζ	ζ	Xi	Ξ	ξ	Chi	Χ	χ
Eta	Η	η	Omicron	Ο	ο	Psi	Ψ	ψ
Theta	Θ	θ	Pi	Π	π	Omega	Ω	ω

AU–18

SPACE HANDBOOK

Prepared by
Air University Institute for Professional Development

AIR UNIVERSITY
MAXWELL AIR FORCE BASE, ALABAMA

Eighth Revision, July 1970

Seventh Revision, July 1969
Sixth Revision, July 1968
Fifth Revision, July 1967
Fourth Revision, July 1966
Third Revision, July 1965
Second Revision, July 1964
First Revision, July 1963
First Edition, December 1962

This publication has been reviewed and approved by competent personnel of the preparing command in accordance with current directives on doctrine, policy, essentiality, propriety, and quality.

ACKNOWLEDGEMENTS

NUMEROUS people have contributed to the original manuscript and annual revision of the Space Handbook. The current edition was prepared by the following members of the faculty, Air University Institute for Professional Development:

EDITOR	*Lt Col Donald A. Brewer*
CHAPTER 1	*Dr. James A. Fraser*
CHAPTER 2	*Maj Donald P. Novak*
	Capt John C. Saffle
CHAPTER 3	*Maj James W. Rivers*
	Maj Ward C. Shaw
CHAPTER 4	*Maj Albert G. Haddad*
CHAPTER 5	*Maj James F. Garber, III*
CHAPTER 6	*Lt Col George G. Deranian*
CHAPTER 7	*Maj Albert G. Haddad*
CHAPTER 8	*Lt Col William S. Van Gilder*
	Maj Bruce M. Parker
CHAPTER 9	*Maj Robert B. Stuart*
CHAPTER 10	*Dr. James A. Fraser*
CHAPTER 11	*Lt Col William F. Goodner*

PREFACE

THIS TEXT was prepared originally and has been revised annually by the Space Directorate of the Air University Institute of Professional Development. The *Space Handbook* serves as the text for the Fundamentals of Space Operations Course, a resident course within AUIPD. As such, the text was written at an intermediate level of academic difficulty but with considerable depth of detail.

The objectives of the Fundamentals of Space Operations Course are:

1. To provide the student with an understanding of the basic physical laws and principles of the space environment, propulsion, orbital mechanics, guidance and control, and atmospheric penetration which permit and limit space operations.

2. To provide the student familiarization with: objectives of the national space effort; current technology; propulsion devices and launch vehicles; electronic applications; present and possible future weapon and support systems including limitations and feasibility.

3. To stimulate thought on new ideas and concepts so that the student may apply more effectively his knowledge in performance of space planning and operational duties.

The Space Handbook also serves as the support text for the Astronautics and Space Operations phase of instruction, Air Force ROTC Aerospace Studies 300 Course.

Recommendations for improvements of *Space Handbook* should be sent to: Commandant, Air University Institute for Professional Development, Maxwell AFB, Alabama, 36112.

CONTENTS

Chapter 3

PROPULSION SYSTEMS

Chapter 4

SPACE VEHICLE ELECTRICAL POWER

Chapter 5

GUIDANCE AND CONTROL

Chapter 6

GLOBAL COMMUNICATIONS

Chapter 7

LASERS

Chapter 8

ATMOSPHERIC PENETRATION

Chapter 9

COMPUTERS

Chapter 10

RELIABILITY OF SPACE SYSTEMS

Chapter 11

BIOASTRONAUTICS

APPENDIXES

CHAPTER 1

THE SPACE ENVIRONMENT

FOR THOUSANDS of years man has looked at the heavens and wondered. What are the stars? What are the planets? The Moon? The shooting stars? Why and how do these heavenly bodies move? Answers derived from superstition, philosophy, religion and fear abound in the literature and folklore of all peoples. Only recently, in the history of man, have answers been found in observation and experimentation. And, even these answers are tentative. The success of the first manned lunar landing, Apollo 11, was truly a milestone in the search for understanding which continues at a quickening pace.

Today there are at least two ways of looking at the space environment. The first is the magnificent look—the look that sees space as the whole universe in terms of both matter and energy. The second is the practical look—the look that sees space as another region in which man has begun travel. The former staggers the imagination and stimulates both wonder and reverence. The latter is the immediate concern of the military man.

This chapter will start with a brief discussion of the universe as it is believed to be today. Following that will be a more detailed presentation of the characteristics of the near-earth space which has immediate military importance.

THE UNIVERSE

Man lives on an Earth which is one of nine planets in separate orbits around a star called the Sun. In addition to the planets, the Solar System contains thirty-two natural satellites, a variable number of artificial satellites, about thirty thousand asteroids, a hundred billion comets and countless specks of dust. These numbers seem impressive. But the Sun, which is the master of the entire system, is more impressive. It contains 99.9 percent of the matter in the Solar System.

The nearest stellar neighbor to the Sun is a star called Alpha Centauri, a conspicuous double star which is visible from the southern hemisphere. About two degrees from it is Proxima Centauri, so named because it was once thought to be even nearer than Alpha Centauri. This group of three stars is about 4.3 light years from the Sun. That is, it takes light, traveling at 186,300 miles per second, about 4.3 years to reach the Solar System. It would take over 100,000 years for a spacecraft travelling at 25,000 miles per hour to make the trip. To date, the fastest man has ever travelled is about 25,000 miles per hour. It is clear that man cannot live long enough to travel to the nearest star. Of course, this is the present

status of such travel. In the future greater speeds, or perhaps the contraction of time which Dr. Albert Einstein predicted, may make such travel a reality.

Another way to visualize these immense distances is to imagine the Sun as represented by a golf ball. On this scale Alpha Centauri would be represented by a pair of golf balls which are a fraction of a mile apart and 500 miles away. Proxima would be a grain of sand about 25 miles from the pair of golf balls. Some of the other closest stars are:

Barnard's Star	6.1 light years
Wolf 359	8 light years
Sirius	8.6 light years
Procyon	11 light years
Vega	27 light years

But these are only the Sun's nearest neighbors. Betelgeuse is 300 light years away. Polaris (the North Star) is 600 light years away. Yet all of these are only a few of a vast array of stars that form a group called the Milky Way.

The name given to a large group of stars, dust, and gas that stay together in a structure is a *galaxy*. The Milky Way is simply the view from Earth of the galaxy in which the Sun is one star. Figure 1 is a view of the Milky Way as it would appear from the side. Figure 2 is its appearance from above. Of course these are artists' conceptions, because such pictures cannot be taken from *inside* the galaxy. The Milky Way is about 100,000 light years in diameter. The Solar System is located in one of the spiral arms of the galaxy, about 30,000 light years from the center.

How many stars are there in this magnificent structure? Of course no one knows the exact answer. However, Dr. Harlow Shapley estimates the number at about 200 billion. All of these are in motion around the center of the galaxy. At the distance of the Solar System from the center of the galaxy, the speed of revolution is about 135 miles per *second* or about 486,000 miles per hour. Even at this tremendous speed, it takes the sun 220 million years to make one trip around the galactic center. Stars closer to the center move faster and those farther from the center move more slowly.

Beyond the Milky Way the telescopes show other objects. What are they? Gas? Dust? Stars? In the year 1755 Immanuel Kant suggested that these were "island universes"—other galaxies similar to the Milky Way, each consisting of billions of stars. However, it was not until 1917 that an astronomer using the Mount Wilson telescope identified a star in one of these objects beyond the Milky Way.

Today we know that these objects are indeed other galaxies. Some have diameters in the order of 7,000 light years. Others have diameters even greater than the Milky Way—about 150,000 light years. The smaller ones probably contain a billion stars. The larger ones may contain two hundred billion or more. How many galaxies are there in the universe? Again the answer must be an estimate, and the most recent one is huge—one hundred billion! Finally, if one

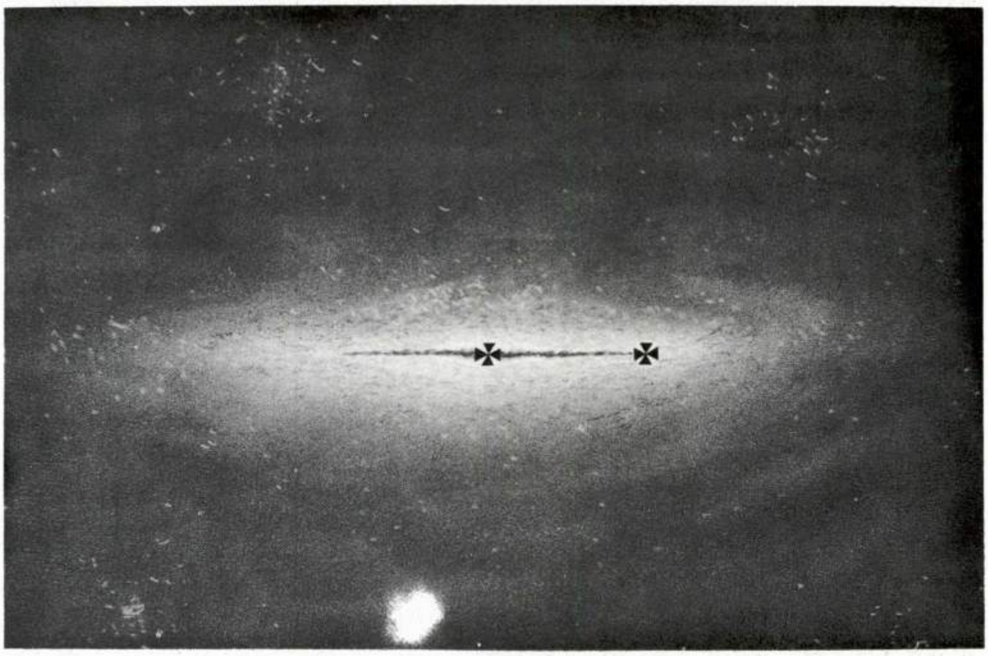


Figure 1

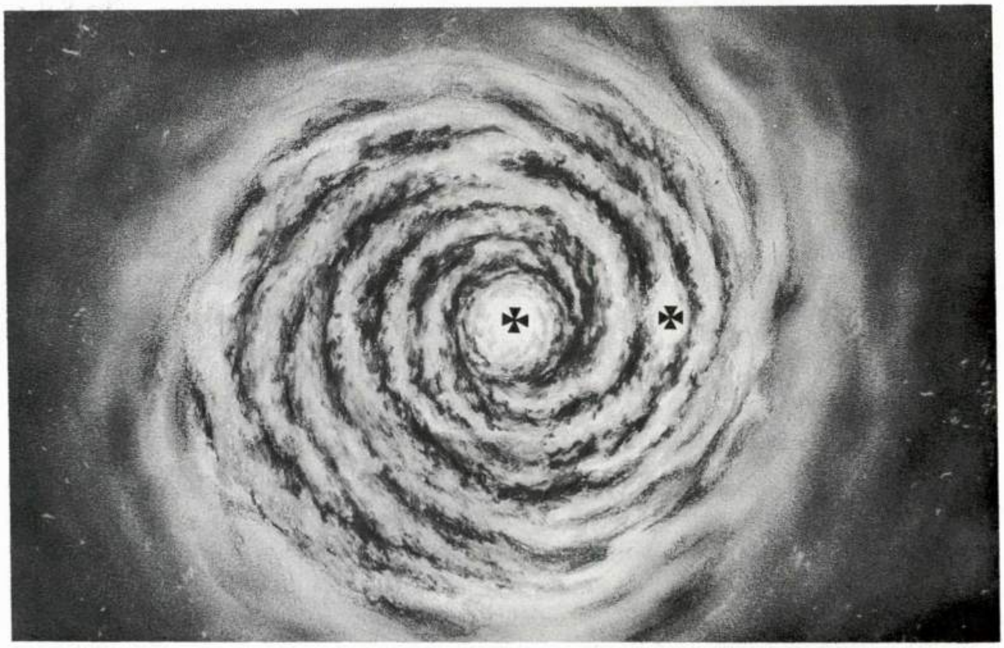


Figure 2

asks the question, "How many stars are there in the Universe?" the answer is approximately 10^{21} stars! Although no one can imagine this number, Sir James Jeans has provided an analogy which helps. He suggests that the number of stars in the Universe is something like *all* the grains of sand on *all* the beaches of *all* the Earth.

To complete the modern idea of this stupendous structure, it is believed that the whole thing is expanding. Every galaxy is rushing away from every other galaxy at tremendous speed! The most distant galaxy visible in a telescope is racing away at about 75,000 miles per second or 270 million miles per hour.

THE MILITARY SPACE ENVIRONMENT

Most of the space discussed above is of little military importance today. However, the near-earth space is important for many military operations. Its characteristics set boundary conditions on both the operations that can be conducted and the equipment required. Consequently, the remainder of this chapter will be devoted to the hazards and physical conditions of near-earth space. "Near-earth" will not be defined precisely, but it is understood to mean not more than 100 million miles from Earth. The Sun is about 93 million miles away. Thus the Sun, the inner planets, and the space between them is the concern of this section. Emphasis will be upon space this side of the Moon.

The first topic to consider is: Where does space begin? Certainly it does not begin at the surface of the Earth because that is where the atmosphere begins. At an altitude of 10,000 feet the oxygen pressure of the atmosphere is not great enough to keep people efficient over a long period of time. Many people, of course, become acclimatized to altitudes of 10,000 feet and higher. But for a man who lives near sea-level, the oxygen pressure at levels above 10,000 feet is insufficient to sustain active and efficient performance. Thus the Air Force requires the use of supplemental oxygen by crew members at altitudes above 10,000 feet.

Approximately one half of the Earth's atmosphere is below an altitude of three miles. But it is not until an altitude of about nine miles that supplemental oxygen fails as a sufficient aid to sustain human life. Here the combined pressure of carbon dioxide and water vapor in the lungs equals the outside atmosphere pressure and breathing cannot take place without supplemental pressure. Hence, at this altitude pressure cabins or pressure suits become a necessity.

The vapor pressure of man's body fluids is about 47 mm of mercury. As soon as the atmospheric pressure drops to this level, bubbles of water vapor and other gases appear in the body fluids. This means literally that the blood will boil. The gas bubbles first appear on the mucous membranes of the mouth and eyes and later in the veins and arteries. This would happen to an unprotected man at an altitude of 12 miles. However, supplemental pressure will suppress this evil.

At 15 miles compressing the outside air to pressurize a cabin is no longer effective. At that altitude the air density is about $\frac{1}{27}$ of the sea level value. Compressing this thin air is perhaps not an impossible task but it certainly is an uneconomical task. Further, the act of compressing air would involve undesirable heat transfer to the air. Finally, at this altitude the atmosphere contains a

significant percentage of ozone. If this were compressed it would poison the cabin atmosphere. Hence, above this altitude the cabin or space suit must have a supply of both oxygen and pressure independent of the outside atmosphere. As far as man is concerned, 15 miles above the surface can be considered the beginning of space. Above this altitude man must take everything he needs with him. His environment will supply him with neither food nor air. He needs a sealed environment containing necessary supplies from Earth.

Five miles farther out, at the 20 mile level, is the operating limit for turbojet engines. At 28 miles ramjet engines do not have enough air to operate. Above this altitude engines must be supplied with *both* a fuel and an oxidizer. Thus, to a propulsion engineer 28 miles above the Earth is the beginning of space. Above this he must use rockets.

In one sense space begins at 50 miles because flight above this altitude earns a crew member the right to wear Astronaut's wings.

In 1964, a New York law firm asked the Air Force Office of Aerospace Research to define the beginning of space. This scientific organization based the answer on aerodynamic forces. Such forces acting on ballistic reentry vehicles, lifting reentry vehicles and boost-glide orbital vehicles can usually be *neglected* at altitudes above 100 kilometers or 62 miles. Thus, for the aeronautical engineer concerned with lift and drag, space may begin at 62 miles.

At about 100 miles above the earth is a region of darkness and utter silence. This is the region of the black sky. The stars appear as brilliant points of light, and between them is absolute black because there is not enough air to scatter light. Neither is there enough air to carry sound or shock waves. There are no sonic booms.

From the above discussion it is clear that there are many answers to the question, "Where does space begin?" The acceptable answer depends upon the reference frame in which the question was asked.

"Is space really nothing?" The answer is "No." Space is filled with surprising amounts of matter and is flooded with energy. First consider the density of matter by starting up from the Earth's surface. At the surface the concentration of particles in air is about a million, million, million particles per cubic centimeter ($10^{18}/cm^3$). There is a decrease in particle density with altitudes and the *average* figures given are only approximate. An average figure for the zone between 7 miles and 50 miles is about $10^{14}/cm^3$. From 50 miles to 600 miles an average figure is about a million particles per cubic centimeter ($10^6/cm^3$). From 600 miles to 1200 miles there are still in the order of 100 particles per cubic centimeter. Above 1200 miles there will be found something like one particle per cubic centimeter. This is certainly far from nothing. There are also localized conditions that cause the particle density to be much higher. Some of these conditions will be discussed later. And, all of space is flooded with electromagnetic energy in many forms from the Sun, from the stars in the Milky Way and even from other galaxies in the Universe.

However, from another point of view, space is "not much." Consider air pressure. At the Earth's surface the pressure varies around the figure 760 mm of mercury. Above 1200 miles the pressure is much less than one mm of mer-

cury. In fact, it probably is around 10^{-12} to 10^{-16} Torr.* This pressure is so low that it is often called a "hard vacuum." It causes some unexpected phenomena. A few of these will be discussed as illustrations.

In the atmosphere any metal is covered with at least a single layer of absorbed gas. In a hard vacuum this film of gas bleeds into space. Metals touching each other tend to weld together. In the atmosphere this doesn't happen because the thin film of air acts as a lubricant keeping the metals apart. To prevent this "cold welding" in space, special measures must be taken.

Some metals are stronger in a hard vacuum. If a crack forms in a metallic surface when the metal is surrounded by air, molecules of air immediately enter the crack. Chemical reaction with the metal occurs. If the reaction product is more voluminous than the original crack, a wedging action occurs and enlarges the crack. In a hard vacuum a chemical reaction causing an enlargement of a crack does not occur. Thus some metals may be stronger in space than they are on Earth.

To study space effects such as the above, it is useful to simulate a hard vacuum on Earth. Late in 1965 USAF completed a large Aerospace Environmental Simulation Chamber at the Arnold Engineering Development Center in Tennessee.

Electromagnetic Radiation

Visible light, ultra-violet, x-rays, infrared, radio and other forms of energy can travel through the hard vacuum of space as electromagnetic radiation. This term refers to the fact that radiation consists of a varying electric field and a varying magnetic field. Together these fields form a wave. Such a wave can be transmitted through a vacuum and does not require the presence of a ma-

---

* A Torr is the same as one millimeter of mercury. It is named in honor of Torricelli and is commonly used in low pressure measurements.

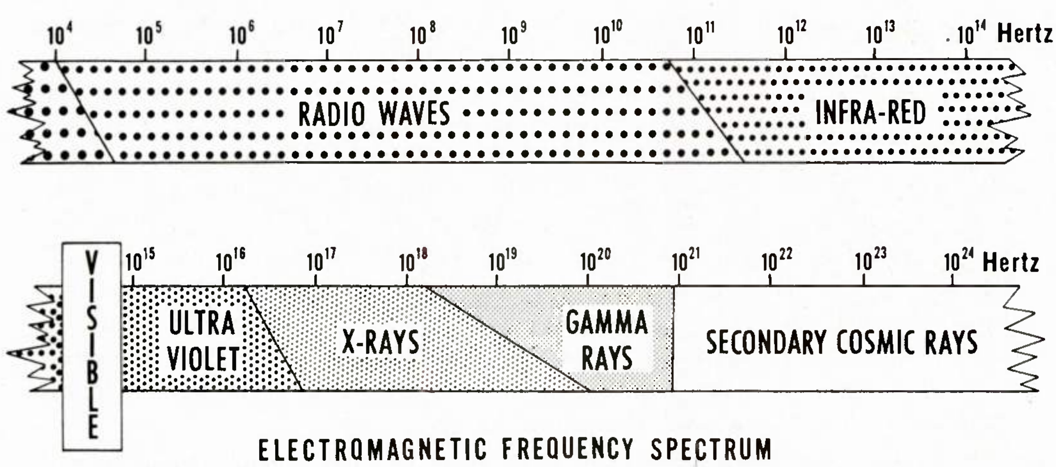


Figure 3

terial medium. This form of energy floods all of space. Its intensity varies with proximity to the Sun, or to a star.

Notice in Figure 3 that visible light covers only a very narrow band in the spectrum of electromagnetic energy. The entire spectrum ranges from frequencies of about 10^4 Hertz up to about 10^{24} Hertz.

Most of the radiation coming from the Sun and stars is absorbed by the Earth's atmosphere before it reaches the surface of the Earth. In fact, there are only two "windows" through which space may be observed. One window includes visible light frequencies and part of the ultraviolet and infrared frequencies. This is called the optical window. Another window called the radio window is found in the radio frequencies of approximately 10^9 cycles per second. Man is protected from (or denied, depending upon the point of view) all other frequencies by the Earth's atmosphere. In space radiation of frequencies throughout the spectrum is present. Man may use it as a source of information, but he must also protect himself against it by a space cabin or space suit.

Meteoroids and Micrometeoroids

The three terms "meteoroid," "meteor," and "meteorite" have similar meaning and are often used interchangeably. *Meteoroid* refers to a particle, large or small, moving in space. When a meteoroid enters the atmosphere and begins to glow, it is then called a *meteor.* If that same particle survives the trip through the atmosphere and hits the earth, the remnant is called a *meteorite.* Some meteoroids must be very large because meteorites with masses of several tons have been found. Most meteoroids, however, are quite small. Extremely small meteoroids are called micrometeoroids.

Meteoroids and micrometeoroids move with speeds varying from about 30,000 miles per hour to 160,000 miles per hour. At these speeds, impact between a satellite and a large meteoroid would be catastrophic. Impacts between micrometeoroids and a satellite would not be catastrophic but could erode the satellite's surface.

Several satellites have rather mysteriously ceased to function, and there is some conjecture that meteoroid impact damage may have been the cause. The probability of meteoroid and micrometoroid impact has been extensively studied. Many methods have been employed. Micrometeoroids so small that it would take about 125 of them to equal the thickness of a piece of paper have been captured, and they have been studied with an electron microscope. The Pegasus satellites, Explorer XXIII and Explorer XVI, are examples of satellites used to study the problem. The Pegasus satellites reported the number of penetrations of their panel materials such as aluminum. Explorer XVI, launched December 16, 1962, reported 62 meteoroid penetrations in 7-½ months of space travel. Mariner IV was hit by over 150 micrometeoroids on its trip to Mars. These and other data have led to the conclusion that the probability of a satellite being hit by very small micrometeoroids is high and of being hit by catastrophically large ones is small.

Figure 4 provides some idea of the probabilities involved in the Apollo Project. The probabilities are based upon an estimated vehicle cross section of

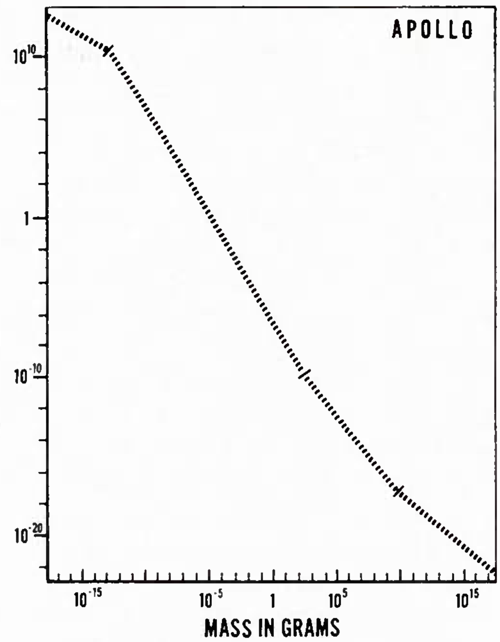


Figure 4

ten square meters and an exposure of ten days. The horizontal axis shows mass of the meteoroids or micrometeoroids in grams. (About 5 grams equals the weight of a nickel.) The vertical axis is scaled in probability until the figure one is reached. This figure means that the vehicle is certain to be hit. Above that, the scale means number of hits in a ten-day period. Notice that the probability of collision with a particle which weighs 10^{-5} grams is 1.0. This kind of particle would penetrate about 0.05 cm into an aluminum skin.

In general, the problem of meteoroid hazard is not as great as was once thought, but it is not negligible. Protection is provided in space suits and capsules. The possibility of a catastrophic hit remains. Extensive studies are continuing.

Cosmic Rays

Cosmic rays are very small particles which have been travelling within the Milky Way for millions of years. Some may even come from other galaxies. They are positively charged and move with great speed. It would take about 10 billion cosmic rays to equal the thickness of a piece of paper. About 84% of them are protons, and about 14% are Alpha particles. The rest are nucleii of atoms with higher atomic numbers ranging from lithium to iron. The speed of cosmic rays is not usually defined. Rather, the kinetic energy level is given. Speed and mass are the two factors which are incorporated into kinetic energy. The energy of cosmic rays is usually stated in electron volts (ev).* In these

* Other frequently used terms are kev (one thousand ev), mev (one million ev) and bev (one billion ev). Energy in electron volts can be converted to kinetic energy. Then if the mass of a proton is known, its speed can be calculated.

units, the energy of cosmic rays varies from about 40 million electron volts to about a *million, million, million* electron volts (10^{12} million ev). Energies greater than a billion electron volts mean that the particle's speed is near that of light.

Cosmic rays arrive at Earth from all directions. Additional protons also arrive at Earth from the direction of the Sun. The former are often called "galactic cosmic rays"; the latter are "solar protons." When either galactic cosmic rays or solar protons collide with atoms of the Earth's atmosphere, the atoms break up, or disintegrate. This process produces a shower of smaller particles which move off in all directions from the site of the collision. The particles are varied in kind and energy but include neutrons, electrons, and mesons, as well as some heavier particles. One of the ways used to study cosmic radiation and its collision with atoms is to fly a photographic plate into space. When the plate is recovered, it frequently will show the result of a cosmic ray striking an atom in the photographic plate and breaking the atom into smaller particles. The charged particles leave traces on the plate; the resulting image resembles a star. Such a collision is often called a "collision star" or "disintegration star."

The Earth's atmosphere serves to protect man from the effects of primary cosmic radiation; in space man would be exposed to a higher intensity of cosmic radiation than he is on Earth. Also, since the cosmic radiation particles are charged particles, the Earth's magnetic field acts as a shield. The effect is to divert many of the particles to the north and to the south. The result is that man on earth is protected by both the Earth's atmosphere and magnetic field. When man ventures into space, he abandons this natural protection.

Beyond the magnetic influence of the Earth, the cosmic radiation from the galaxy presents a flux of one to two particles per square centimeter per second. Despite the high energy of galactic cosmic rays, there will be little hazard to man because of the low flux. Even in a year's time the total radiation dose expected from this source is from 6 to 20 rads. This is less than one-tenth of the dose that would make one-half of recipient humans feel sick or nauseated.

Solar Flares

The high speed solar protons emitted by a solar flare are probably the most potent of the radiation hazards to space flight. Flares themselves are probably the most spectacular disturbances seen on the sun. They are observed optically as a sudden, large increase in light from a portion of the Sun's atmosphere. A flare may spread in area during its lifetime which may be from several minutes to a few hours. Flares are classified according to a range of importance of 0 to 4.

There is a relationship between the number of sunspots and the frequency of flare formation, but the most important flares do not necessarily occur at sunspot maximum.

There are many events that may occur at Earth following a solar flare although not many flares produce *all* of the possible events. In addition to the increase in visible light, minutes after the start of a flare there is a Sudden Ionospheric Disturbance (SID) in the Earth's ionosphere. This, in turn, causes short wave fade-out, resulting in the loss of long range communications for 15 minutes to 1 hour. X-rays emitted by the flare probably cause the SID's. During the first few minutes

TABLE 1
Solar Flare Classification

Importance	Area* (Millionths of Solar Hemisphere)	Average Duration
0	Less than 100	17 min
1	100 to 249	32 min
2	250 to 599	69 min
3	600 to 1200	145 min
4	Greater than 1200	145 min
BRIGHTNESS CATEGORIES	FAINT (F) NORMAL (N) BRILLIANT (B)	

* The area of the earth's disk is approximately equivalent to the area of an Importance one flare.

of a flare there may be a radio noise storm, consisting of bursts of noise over a wide range of frequencies. In addition, there may be disturbances in the Earth's magnetic field, changes in the Auroras, and *decreases* in galactic cosmic ray intensity. However, from the point of view of a space traveller, by far the most important effect is the marked increase in solar protons. The energy of these protons ranges from about 10 million electron volts to about 500 million electron volts. The flux may be quite high. Consequently, the dose of radiation accumulated during exposure to the solar protons may vary from negligible to well above a lethal dose.

Construction of space vehicles sufficiently shielded to protect against all possible solar proton events is impractical due to the amount and density of required shielding materials. Consequently the best hope for protection lies in developing a method for reliable prediction of flare occurrence and intensity. The USAF Air Weather Service is charged with this responsibility.

The Van Allen Radiation Belt

Another problem that man must overcome in venturing into space is trapped radiation. As a result of experiments conducted in 1958 with the US Explorer I satellite and subsequent experiments, Dr. James A. Van Allen and his associates discovered the existence of geomagnetically trapped particles encircling the earth. When electrons and protons, and perhaps some other charged particles, encounter the Earth's magnetic field, many of them are trapped by the field. They oscillate back and forth along the lines of force, and since the magnetic field completely encircles the Earth, the trapped particles completely encircle the Earth.

The belt has an inner and outer portion. Recent data shows that the toroidal shaped volume occupied by the Van Allen radiation is permeated with both protons and electrons. The protons are most intense at about 2,200 miles. The electron flux peaks at about 9,900 miles. The low particle density separating the two belts is often called the "slot." In this particular volume of space some phenomena, as yet not fully undestood, reduces the lifetime of the charged particles.

The inner Van Allen belt starts at an altitude of about 250 miles to 750 miles, depending upon latitude. It extends to about 6,200 miles where it begins to over-

lap the outer belt and where the "slot" begins. The inner belt extends from about 45° north latitude to about 45° south latitude.

The outer Van Allen belt begins at about 6,200 miles and extends to between 37,000 and 52,000 miles. The upper boundary is dependent upon the activity of the Sun.

Both the inner and outer belts were affected by a high altitude nuclear device test in July 1962. Radiation in both belts increased after the detonation, and the low radiation slot separating the two belts was eliminated for some time.

Experience has shown that space vehicles in low circular orbit (125-350 miles) receive an insignificant amount of radiation from the Van Allen zones. However, a vehicle in a highly eccentric orbit or one in a high altitude circular orbit can receive an important dose. For example, a satellite in a synchronous orbit over the equator will be close enough to the center of the outer zone to accumulate a hazardous dose. But, as demonstrated by Apollo lunar missions, man can safely transit these zones in a spacecraft with minimal shielding by judicious selection of the flight trajectory.

The Van Allen belt varies daily with changes in the magnetosphere. On the Sun side of Earth it is flattened. On the night side of the Earth it is elongated.

The Solar Wind

Because of the high temperature of the Sun's corona, protons and electrons beyond a certain distance from the Sun acquire velocities in excess of the escape velocity from the Sun. Thus there is a continuous outward flow of charged particles in all directions from the Sun. This has been called the solar wind. It is a plasma wind, rather than a gas wind. Its velocity and density vary with sunspot activity. During the time of sunspot minimum at the Earth's distance from the Sun, the density is about 100 particles per cubic centimeter. The speed is about 300 miles per second. At sunspot maximum the corresponding density is probably around 10,000 particles per cubic centimeter, and the speed is about 900 miles per second. When the solar wind encounters the Earth's magnetosphere, it flows around the magnetosphere, which is flattened in the process. On the dark side of the Earth the magnetosphere is elongated.

The energy of the particles in the solar wind is not high. No hazard to man is expected from the wind. However, it is either the cause of or a contributor to the aurora which illuminates the polar night sky. Within the past few years ground based observations have been combined with information acquired by rockets and earth satellites to provide an explanation of this previously baffling beauty. The magnetosphere of the Earth acts like a giant cathode-ray tube, directing charged particles from the solar wind into beams and focusing them on the Earth's polar regions. The aurora is a fluorescent luminosity produced by the electrons and protons of the solar wind which strike atoms and molecules of oxygen and nitrogen high in the atmosphere of the polar region.

The radiation hazard of space may be summarized as follows:

(1) Electromagnetic energy in space may be shielded and is not a serious threat to life.

(2) Electromagnetic energy may upset radio communication and guidance equipment.

(3) Galactic cosmic rays and solar wind do not present a serious threat to space travel.

(4) Van Allen radiation does present a serious threat, but the location of the belts is sufficiently well known that flight trajectories may be planned to limit time spent in the hazardous regions.

(5) Protons emitted at the time of a solar flare present the greatest uncertainty and the greatest threat to manned flight in regions beyond the protection of the Earth's atmosphere and magnetosphere.

BIBLIOGRAPHY

Air Force Cambridge Research Laboratories. *AFCRL in Space.* L. G. Hanscom Field, Bedford, Mass: Office of Aerospace Research, 1967.

Cosby, William A. and Lyle, Robert G. *The Meteoroid Environment and its Effects on Materials and Equipment.* Washington, DC: U.S. Government Printing Office, 1965.

Guide to Solar-Geophysical Activity, 4th Weather Wing Pamphlet, 4WWP 105–1, Department of the Air Force, Headquarters 4th Weather Wing (MAC), Ent Air Force Base, Colorado, 1 February 1969.

Flinders, Col. Dale J. "The Space Forecasting System: Confluence of Military and Scientific Interests," *Air University Review,* Nov-Dec 1969, pp. 37–50.

Glasstone, Samuel. *Sourcebook on The Space Sciences.* New York: D. Van Nostrand Company, Inc., 1965. Chapters 6 through 12.

Forecasting Solar Activity and Geophysical Responses, 4th Weather Wing Manual, 4WWM 105–1, Solar Forecast Facility Headquarters, 4th Weather Wing, Ent Air Force Base, Colorado, 28 June 1968.

Office of Aerospace Research. "The Videometer, A New Instrument for Solar Flare Analysis," *OAR Research Review,* Vol. IX No. 1, 1970, p. 7.

Zimmerman, Frank J. *Meteoroid Threat to Extravehicular Space Suit Assemblies.* Aerospace Medical Research Laboratory, AFSC, Wright-Patterson AFB, AMRL-TR-68-86, June 1969.

CHAPTER 2

ORBITAL MECHANICS

THE STUDY of trajectories and orbits of vehicles in space is not a new science but is the application of the concepts of celestial mechanics to space vehicles. Celestial mechanics, which is mainly concerned with the determination of trajectories and orbits in space, has been of interest to man for a long time. When the orbiting bodies are man-made (rather than celestial), the topic is generally known as orbital mechanics.

The early Greeks postulated a fixed earth with the planets and other celestial bodies moving around the earth, a geocentric universe. About 300 B. C., Aristarchus of Samos suggested that the sun was fixed and that the planets, including the earth, were in circular orbits around the sun. Because Aristarchus' ideas were too revolutionary for his day and age, they were rejected, and the geocentric theory continued to be the accepted theory. In the second century A.D., Ptolemy amplified the geocentric theory by explaining the apparent motion of the planets by a "wheel inside a wheel" arrangement. According to this theory, the planets revolve about imaginary planets, which in turn revolve around the earth. It is surprising to note that, even though Ptolemy considered the system as geocentric, his calculations of the distance to the moon were in error by only 2%. Finally, in the year 1543, some 1800 years after Aristarchus had proposed a heliocentric (sun-centered) system, a Polish monk named Copernicus published his *De Revolutionibus Orbium Coelestium,* which again proposed the heliocentric theory. This work represented an advance, but there were still some inaccuracies in the theory. For example, Copernicus thought that the orbital paths of all planets were circles and that the centers of the circles were displaced from the center of the sun.

The next step in the field of celestial mechanics was a giant one made by a German astronomer, Johannes Kepler (1571–1630). After analyzing the data from his own observations and those of the Danish astronomer Tycho Brahe, Kepler stated his three laws of planetary motion.

A contemporary of Kepler's, named Galileo, proposed some new ideas and conducted experiments, the results of which finally caused acceptance of the heliocentric theory. Some of Galileo's ideas were expanded and improved by Newton and became the foundation for Newton's three laws of motion. Newton's laws of motion, with his law of universal gravitation, made it possible to prove mathematically that Kepler's laws of planetary motion are valid.

Kepler's and Newton's work brought celestial mechanics to its modern state of development, and the major improvements since the days of Newton have been mainly in mathematical techniques, which make orbital calculations easier.

Because the computation of orbits and trajectories is the basis for predicting and controlling the motion of all bodies in space, this chapter describes the fundamental principles of orbital mechanics upon which these computations are based. It also shows how these principles apply to the orbits and trajectories used in space operations.

MOTION OF BODIES IN ORBIT

Bodies in space move in accordance with defined physical laws. Analysis of orbital paths is accomplished by applying these laws to specific cases. Orbital motion is different from motion on the surface of the earth; however, many concepts and terms are transferable, and similar logic can be applied in both cases. An understanding of simplified linear and angular motion will permit a more thorough appreciation of a satellite's path in space.

Linear Motion

Bodies in space are observed to be continuously in motion because they are in different positions at different times. In describing motion, it is important to use a reference system. Otherwise, misunderstanding and inaccuracies are likely to result. For example, a passenger on an airliner may say that the stewardess moves up the aisle at a rate of about 5 ft per sec, but, to the man on the ground, the stewardess moves at a rate of the aircraft's velocity plus 5 ft per sec. The man in the air and the man on the ground are not using the same reference system. For the present, the matter of a reference system will be simplified by first describing movement along a straight line, or what is called rectilinear motion.

Rectilinear motion can be described in terms of speed, time, and distance. Speed is the distance traveled in a unit of time, or the time rate of change of distance. An object has uniform speed when it moves over equal distances in equal periods of time. Speed does not, however, completely describe motion.

Motion is more adequately described if a direction as well as a speed is given. A speedometer tells how fast an automobile is going. If a direction is associated with speed, the motion is now described as a velocity. A velocity has both a magnitude (speed) and a direction, and it is therefore a vector quantity.

Uniform speed in a straight line is not the same as uniform speed along a curve. If a body has uniform motion along a straight line for a given time, then average velocity is represented by the equation $v = \frac{s_f - s_o}{t_f - t_o}$. In the equation, s_o is the initial position, s_f is the final position, t_o is the initial time, and t_f is the final time; or more simply, the velocity is the change in position divided by the change in time. The units of velocity are distance divided by time, such as ft per sec or knots (nautical miles per hour).* Since velocity is a vector quantity, it may be treated mathematically or graphically as a vector.

If velocity is not constant from point to point (i.e., if either direction or speed is changed), there is acceleration. Acceleration, which is also a vector quantity,

* The nautical mile (NM) is one minute of a great circle. In this course, use the conversion that 1 NM = 6,080 feet = 1.15 statute miles.

is the time rate of change of velocity. The simplest type of acceleration is one in which the motion is always in the same direction and the velocity changes equal amounts in equal lengths of time. If this occurs, the acceleration is constant, and the motion can be described as being uniformly accelerated.

The equation $a_{av} = \frac{v_f - v_o}{t_f - t_o}$ defines the average acceleration, over the specified time interval. A good example of a constant acceleration is that of a free-falling body in a vacuum near the surface of the earth. This acceleration has been measured as approximately 32.2 ft per sec per sec, or 32.2 ft per sec^2. It is usually given the symbol g. Since an acceleration is a change in velocity over a period of time, its units are ft/sec^2, or more generally, a length over a time squared. Actually, constant acceleration rarely exists, but the concepts of constant acceleration can be adapted to situations where the acceleration is not constant.

The following three equations are useful in the solutions of problems involving linear motion:

$$(1)\ s = v_o t + \frac{at^2}{2}$$

$$(2)\ v_f = v_o + at$$

$$(3)\ 2as = v_f^2 - v_o^2$$

where s is linear displacement, v_o is initial linear velocity, v_f is final linear velocity, a is constant linear acceleration, and t is the time interval.

Angular Motion

If a particle moves along the circumference of a circle with a constant tangential speed, the particle is in uniform circular motion. Since velocity signifies both speed and direction, however, the velocity is constantly changing because the direction of motion is constantly changing. Now, acceleration is defined as the time rate of change of velocity. Since the velocity in uniform circular motion is changing, there must be an acceleration. If this acceleration acted in the direction of motion, that is, the tangential direction, the magnitude of the velocity (the speed) would change. But, since the original statement assumed that the speed was constant, the acceleration in the tangential direction must be equal to zero. Therefore, any acceleration that exists must be perpendicular to the tangential direction, or in other words, any acceleration must be in the radial direction (along the radius).

Average speed is equal to the distance traveled divided by the elapsed time. For uniform circular motion, the distance in one lap around the circle is $2\pi r$, which is covered in one period (P). Period is the time required to make one trip around the circumference of the circle. Therefore, the tangential speed $v_t = \frac{2\pi r}{P}$. In uniform circular motion, the particle stays the same distance from the center, therefore, radial speed, $v_r = 0$. It has already been shown that $a_t = 0$; and it will be shown in the next section that $a_r = \frac{4\pi^2 r}{P^2} = \frac{v_t^2}{r}$.

Translatory motion is concerned with linear displacement, s; velocity, v; and acceleration, a. Angular motion uses an analogous set of quantities called angular displacement, θ; angular velocity, ω; and angular acceleration, α.

In describing angular motion, it is convenient to think of it in terms of the rotation of a radius arm (r), as shown in Figure 1. The radius arm initially coincided with the polar axis, but at some time later (t seconds) it was positioned as shown.

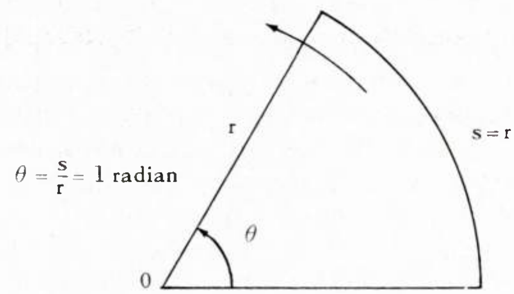


Figure 1. Position of radius arm as rotated one radian (57.3 degrees) from the starting point.

Angular displacement (θ) is measured in degrees or radians. A radian is the angle at the center of a circular arc which subtends an arc length equal to the radius length. If the length of s equaled the length of r, θ would be equal to one radian, or 57.3°. The central angle of a complete circle is 360° or 2π radians ($2\pi = 6.28$).

The following equations for angular motion are analogous to those studied earlier for rectilinear motion:

$$\theta = \frac{s}{r} \text{ radians}$$

$$\omega_{av} = \frac{\theta_f - \theta_o}{t_f - t_o} \text{ rad/sec}$$

$$\alpha_{av} = \frac{\omega_f - \omega_o}{t_f - t_o} \text{ rad/sec}^2$$

$$\omega_f = \omega_o + \alpha t$$

$$\theta = \omega_o + \frac{\alpha t^2}{2}$$

$$2\alpha\theta = \omega_f^2 - \omega_o^2$$

In the equations, θ_f is final angular position; θ_o is initial angular position; s is linear displacement (arc length); r is the radius; ω is average angular speed; ω_f is final angular speed; ω_o is initial angular speed; t_f is final time; t_o is initial time; and α is constant angular acceleration.

If a body is rotating about a center on a radius *r*, the tangential linear quantities are related to the angular quantities by the following formulas [where θ, ω, and α are in radians]:

$$s = r\theta$$

$$v_t = r\omega$$

$$a_t = r\alpha$$

Principles of the Calculus Applied to Astronautics

Computations in the calculus are based upon the idea of a limit of a variable. According to the formal definition, *the variable x is said to approach the constant 1 as a limit when the successive values of x are such that the absolute value of the difference x − 1 ultimately becomes and remains less than any preassigned positive number, however small.*

An example will make the definition easier to understand. The area of a regular polygon inscribed in a circle approaches the area of the circle as a limit as the number of sides of the polygon approaches infinity (Fig. 2).

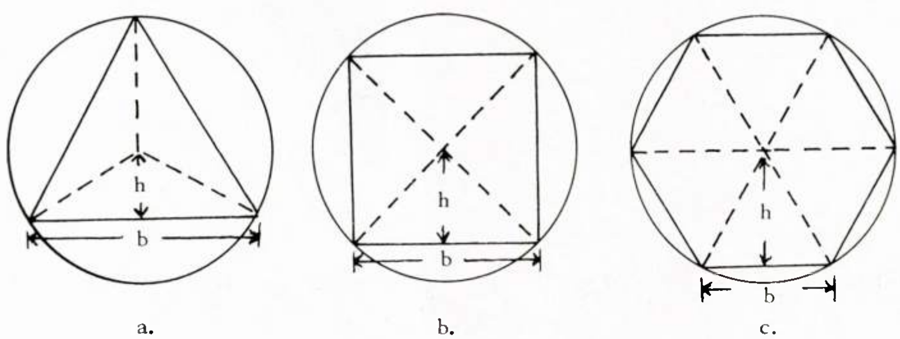


Figure 2. Increase in the number of sides of a regular polygon inscribed in a circle.

The area of a triangle is 1/2 bh. In general, if there are n sides to a polygon, the polygon is made up of n triangles as shown in Figure 2. Therefore, the area of the polygon is 1/2 nbh. As the number of sides (n) approaches infinity as a limit, the product nb approaches the circumference of the circle (c). Also, as n approaches infinity, the value of h approaches the radius (r) as a limit.

$$\lim_{n \to \infty} 1/2 \text{ nbh} = \frac{cr}{2}$$

This is read, "The limit of 1/2 nbh as n approaches infinity is equal to $\frac{cr}{2}$."

$$\text{But } c = 2\pi r$$

$$\therefore \lim_{n \to \infty} \text{area of the polygon} = \lim_{n \to \infty} \frac{nbh}{2} = \frac{(2\pi r)r}{2}$$

$$\text{and } \frac{(2\pi r)r}{2} = \pi r^2 = \text{area of the circle}$$

Now, an increment is the difference in two values of a variable. In the example above, the increase in area when the inscribed polygon increases the number of sides by one is an increment of area; that is, the area of an inscribed square minus the area of an inscribed triangle is an increment of area. An increment is written as Δx which is read "delta x," and does not mean Δ multiplied by x.

In the previous section the radial acceleration for uniform circular motion was given as $a_r = \frac{v_t^2}{r}$. With the concepts of an increment and a limit, the value for radial acceleration can be determined mathematically. In Figure 3, an increment of arc has been expanded to permit closer examination. The length r is the distance from the center of the circle to the circumference. The horizontal distance $v_t \Delta t$ is

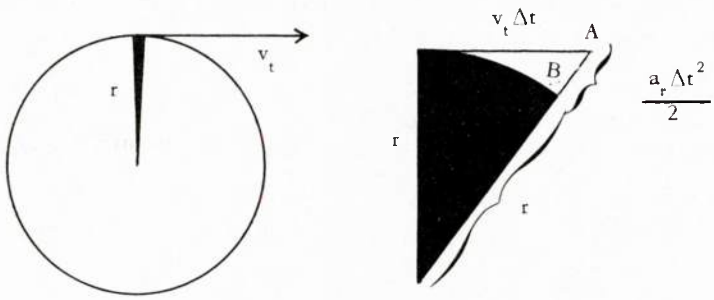


Figure 3. An increment of an arc (left) and the increment expanded (right) to show change in velocity.

the distance a body in uniform motion with a velocity v_t would move in the time Δt. However, at the completion of the increment of time the body is not at point A but at point B, because this is uniform circular motion. The distance from A to B is equal to $v_r \Delta t + \frac{a_r \Delta t^2}{2}$ where the subscript r refers to radial. However, for uniform circular motion $v_r = 0$. Therefore, the distance AB is equal to $\frac{a_r \Delta t^2}{2}$. Now, applying the Pythagorean theorem to the triangle,

$$r^2 + (v_t \Delta t)^2 = \left[r + \frac{a_r \Delta t^2}{2} \right]^2$$

$$\text{or, } r^2 + v_t^2 \Delta t^2 = r^2 + r a_r \Delta t^2 + \frac{a_r^2 \Delta t^4}{4}$$

Subtracting r^2 from both sides,

$$v_t^2 \Delta t^2 = r a_r \Delta t^2 + \frac{a_r^2 \Delta t^4}{4}$$

Dividing both sides by Δt^2

$$v_t^2 = r a_r + \frac{a_r^2 \Delta t^2}{4}$$

To find the instantaneous values take the limit as $\Delta t \to 0$.

$$v_t^2 = r\ a_r + \frac{a_r^2\ (0)}{4} = r\ a_r$$

$$\therefore\ a_r = \frac{v_t^2}{r} \quad \text{As was to be demonstrated.}$$

This text does not attempt to teach the processes of differentiating and integrating, but its purpose is to give the student some understanding of how the calculus is used in the study of space.

The definition of the derivative of y with respect to x is, in symbol form, $\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$. Any calculus book has a table of derivatives, and there is also one in *The Engineer's Manual* by Ralph G. Hudson on pages 31 and 32.

The average velocity over a period of time, as given in the previous section, is:

$$v = \frac{s_f - s_o}{t_f - t_o}$$

Usually the average velocity is not of direct value in analysis, but the instantaneous velocity is. The speedometer in a car measures instantaneous speed, and if a motorist is arrested for speeding, it is because of his instantaneous velocity, not his average velocity. If s is the path of a particle, its instantaneous velocity is equal to: $\frac{ds}{dt} = \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t}$.

Example: A particle moves so that its distance from the origin at any time follows the formula $s = t^3$. Find its average and final, velocity and acceleration after 3 seconds.

$$v = \frac{s_f - s_o}{t_f - t_o} \qquad t_o = 0, \quad s_o = 0$$

$$t_f = 3, \quad s_f = 27$$

$$v_{av} = \frac{27 - 0}{3 - 0} = 9 \quad \text{Answer}$$

$$v_f = \frac{ds}{dt} = \frac{d}{dt}(t^3)$$

From page 32 of *The Engineer's Manual:*

$$\frac{d}{dx}(u^n) = n\ u^{n-1} \frac{du}{dx}$$

$$\frac{d}{dt}(t^3) = 3t^{3-1} \frac{dt}{dt} = 3t^2$$

$$v_f = 3(3)^2 = 27 \quad \text{Answer}$$

$$a_{av} = \frac{v_f - v_o}{t_f - t_o} = \frac{27 - 0}{3} = 9 \quad \text{Answer}$$

$$a_f = \frac{dv_f}{dt} = \frac{d(3t^2)}{dt} = 2\ (3t^{2-1})\frac{dt}{dt} = 6t = 18 \quad \text{Answer}$$

Note that with the use of differential calculus, final or instantaneous values for velocity and acceleration can be determined, but only average values can be determined from the formulas given in the previous section.

If, in the example above, the acceleration were given as $a_f = 6t$, the instantaneous velocity and position could be determined by the process of integration. Integral calculus is a summation process that is the inverse of differential calculus.

Example: $a_f = 6t$. Find v_f after 3 seconds. If a curve is drawn with acceleration on the vertical axis and time on the horizontal axis, the area under the curve is the velocity (Fig. 4). Integration gives the sum of all the individual shaded rectangles as

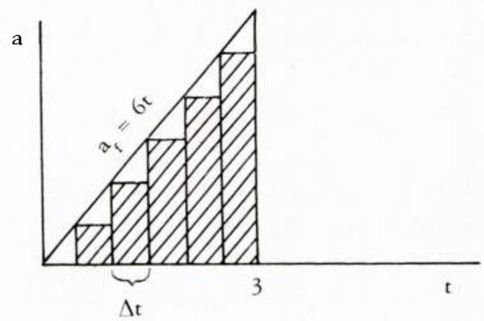


Figure 4. Graph of the continuous function $a_f = 6t$.

Δ t approaches 0 as a limit. As $\Delta t \to 0$, the area of the rectangles approaches the area under the curve as a limit and is the velocity in this problem. The symbol for integration is $\int$. From the table of fundamental theorems on integrals (Hudson's *The Engineer's Manual,* p. 39),

$$\int u^n du = \frac{u^{n+1}}{n+1} + c$$

In this example problem, the limits of integration, t = 0 to t = 3, are specified, so the + c (constant of integration) may be dropped.

$$v_f = \int_0^3 a_f dt = \int_0^3 6t\ dt = \frac{6t^{1+1}}{2} \Bigg|_0^3$$

$$v_f = 3t^2 \Big|_0^3 = 3\,(3)^2 - 3\,(0)^2 = 27 \qquad \text{Answer}$$

The processes of integration and differentiation of variables as applied to the computation of velocity and acceleration through the calculus are part of the study of motion taken up in the branch of dynamics known as kinematics. The study of the forces causing the motion belongs to another branch of dynamics called kinetics.

LAWS OF MOTION

Natural bodies in space follow the basic laws of dynamics, as described by Newton's universal law of gravitation and his three laws of motion. By applying the basic laws and making use of calculus (also developed by Newton), one can explain and prove Kepler's three laws of planetary motion. It would be well to review Kepler's laws before stating Newton's law of universal gravitation, which is one of the laws upon which computation of trajectories and orbits* is based, and Newton's three laws of motion, which describe terrestrial motion as well as celestial mechanics.

Kepler's Laws

From his observations and study, Kepler concluded that the planets travel around the sun in an orbit that is not quite circular. He stated his first law thus: *The orbit of each planet is an ellipse with the sun at one focus.*

Later Newton found that certain refinements had to be made to Kepler's first law to take into·account perturbing influences. As the law is applied to manmade satellites, we must assume that perturbing influences like air resistance, the non-spherical (pear shape) shape of the earth, and the influence of other heavenly bodies are negligible. The law as applied to statellites is as follows: *The orbit of a satellite is an ellipse with the center of the earth at one focus.* The path of a ballistic missile, not including the powered and reentry portions, is also an ellipse, but one that happens to intersect the surface of the earth.

Kepler's second law, or law of areas, states: *Every planet revolves so that the line joining it to the center of the sun sweeps over equal areas in equal times.*

To fit earth orbital systems, the law should be restated thus: *Every satellite orbits so that the line joining it with the center of the earth sweeps over equal areas in equal time intervals.*

When the orbit is circular, the application of Kepler's second law is clear, as shown in Figure 5. In making one complete revolution in a circular orbit, a satellite at a constant distance from the center of the earth (radius r) would, for example, sweep out eight equal areas in the total time period (P = 1). Each of these eight areas is equal and symmetrical. According to Kepler's second law, the time required to sweep out each of the eight areas is the same. When a satellite is traversing a circular orbit, therefore, its speed is constant.

When the orbit is elliptical rather than circular, the application of Kepler's second law is not quite so easy to see; although the areas are equal, they are not symmetrical (Fig. 6). Note, for example, that the arc of Sector I is much longer than the arc of Sector V. Therefore, since the radius vector sweeps equal

* The terms "trajectory" and "orbit" are sometimes used interchangeably. Use of the term "trajectory" came to astronautics from ballistics, the science of the motion of projectiles shot from artillery or firearms, or of bombs dropped from aircraft. The term "orbit" is used in referring to natural bodies, spacecraft, and manmade satellites. It is the path made by a body in its revolution about another body, as by a planet about the sun or by an artificial satellite about the earth.

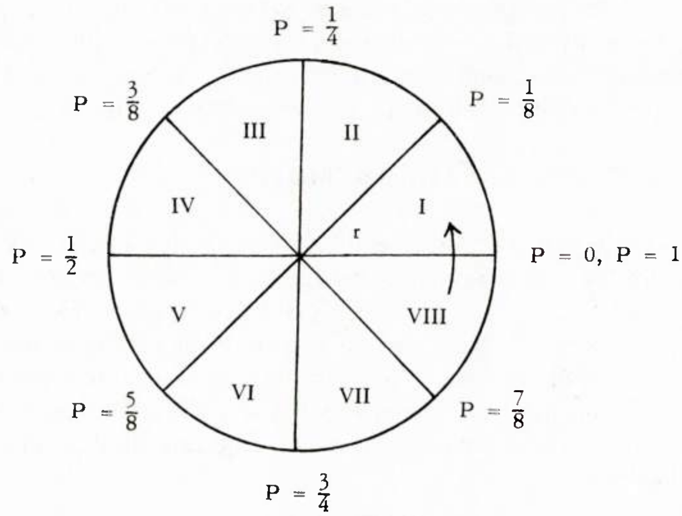


Figure 5. Law of areas as applied to a circular orbit.

areas in equal fractions of the total time period, the satellite must travel much faster around Sector I (near perigee) than around Sector V (near apogee). The perigee (a word derived from the Greek prefix *peri-,* meaning "near," and the Greek root *ge,* meaning "pertaining to the earth") is the point of the orbit nearest the earth. The apogee is that point in the orbit at the greatest distance from the earth (the Greek prefix *apo-* means "from" or "away from").

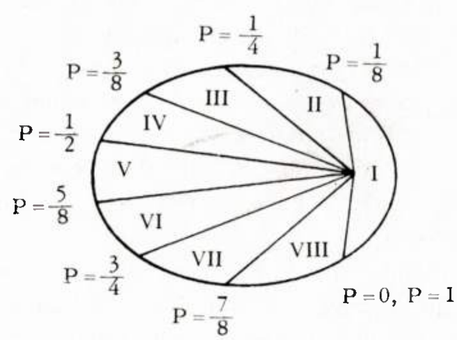


Figure 6. Law of areas as applied to an elliptical orbit.

Kepler's third law, also known as the harmonic law, states: *The squares of the sidereal periods* of any two planets are to each other as the cubes of their mean distances from the center of the sun.*

To fit an earth orbital system, Kepler's third law should be restated thus: *The squares of the periods of the orbits of two satellites are proportional to each other as the cubes of their mean distances from the center of the earth.* The

* The period of a planet about the sun.

mean distance is the length of the semimajor axis (a) of the ellipse, which is an average of the distances to perigee and apogee. Of course, in a circular orbit, the mean distance is the radius, r.

Newton's Laws

While Kepler was working out his three laws of planetary motion, Galileo, an Italian physicist and astronomer, was studying the effects of gravity on falling bodies. Newton drew upon the work of both Kepler and Galileo to formulate his laws of motion.

Newton's first law states: *Every body continues in a state of rest or of uniform motion in a straight line, unless it is compelled to change that state by a force imposed upon it.* In other words, a body at rest tends to remain at rest, and a body in motion tends to remain in motion unless it is acted upon by an outside force. This law is sometimes referred to as the law of inertia.

The second law of motion as stated by Newton says: *When a force is applied to a body, the time rate of change of momentum is proportional to, and in the direction of, the applied force.* If the mass remains constant, this law can be written as $F = Ma$.

Newton's third law of motion is the law of action and reaction: *For every action there is a reaction that is equal in magnitude but opposite in direction to the action.* If body A exerts a force on body B, then body B exerts an equal force in the opposite direction on body A.

Force as Measured in the English System

Newton's three laws of motion are stated in terms of four quantities: force, mass, length, and time. Three of these, length, time and either force or mass, may be completely independent, and the fourth is defined in terms of the other three by Newton's Second Law. Since the units and relative values of these quantities were not known, Newton stated his second law as a proportionality. Assuming that mass does not change with time, this proportionality is stated as $F \propto ma$. If proper units are selected, this statement may be written as an equation:

$$F = ma$$

The following are used in the *metric* system of measurement:

F (dynes) = m (grams) times a (centimeters per second per second)

F (Newtons) = m (kilograms) times a (meters per second per second)

The most common force experienced is that of weight, the measure of the body's gravitational attraction to the earth or other spatial body. Since this attraction is toward the center of the earth, weight, like any force, is a vector quantity. When the only force concerned is weight, the resulting acceleration is

normally called "g," the acceleration due to gravity. For this special case, Newton's Second Law can then be written:

$$W = Mg$$

This equation can be used as a definition of mass. The value of g near the surface of the earth is approximately 32.2 feet per second per second; "g" is a vector quantity since it is directed always toward the center of the earth. If the weight, W, is expressed in pounds, rearranging gives:

$$M = \frac{W \text{ (pounds)}}{g \text{ (feet/sec}^2)}$$

The unit of mass in this equation is called a "slug." Note that mass is a scalar quantity* and is an inherent property of the amount of matter in a body. Mass is independent of the gravitational field, whereas weight is dependent upon the field, the position in the field, and the mass of the body being weighed.

Finally, Newton's Second Law may now be written:

$$F \text{ (pounds)} = M \text{ (slugs) times } a \text{ (ft/sec}^2)$$

The following example shows the use of this system of units and the magnitude of the "slug":

A package on earth weighs 161 pounds.

Find: (a) its mass in slugs.

(b) the force necessary to just lift it vertically from a surface.

(c) the force necessary to accelerate it 10 ft/sec² on a smooth, level surface.

(d) its weight if it were on the moon; assume the value of "g" there is ⅙ of that value here on the earth.

(e) its mass on the moon.

Solution:

(a) $M \text{ (slugs)} = \frac{W}{g} = \frac{161 \text{ pounds}}{32.2 \text{ ft/sec}^2} = 5 \text{ slugs}$

(b) $F = W = M\,g = (5 \text{ slugs})\ (32.2 \text{ ft/sec}^2) = 161 \text{ pounds}$
The force must be applied upwards, in the direction opposite to weight.

(c) $F = M\,a = (5 \text{ slugs})\ (10 \text{ ft/sec}^2) = 50 \text{ pounds}$

(d) $W_{moon} = M\,g_{moon} = (5 \text{ slugs})\ (\frac{32.2}{6} \text{ ft/sec}^2) = 26.83 \text{ pounds}$

(e) $M_{moon} = \frac{W_{moon}}{g_{moon}} = \frac{26.83 \text{ pounds}}{\frac{32.2}{6} \text{ ft/sec}^2} = 5 \text{ slugs}$

* A scalar quantity has magnitude *only*, in contrast to a vector quantity which has magnitude and direction.

This last solution is, of course, to reemphasize that mass is independent of position. It will be shown later that the local value of g varies with altitude above the earth. It is significant to note that the weight will vary such that the ratio $\frac{W}{g}$ remains constant.

Energy and Work

Work, w, is defined as the product of the component of force in the direction of motion and the distance moved. Thus, if a force, F, is applied and a body moves a distance, s, in the direction the force is applied, w = Fs (Fig. 7). The units of work are foot-pounds. Work is a scalar as distinguished from a vector quantity.

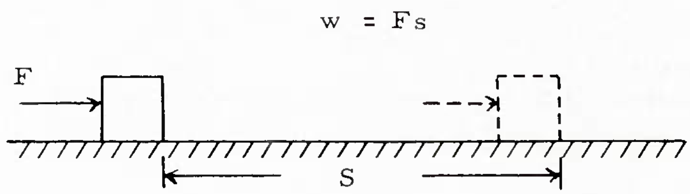


Figure 7. Work performed as a force (F) is moved over the distance s.

To do work against gravity, a force must be applied to overcome the weight, which is the force caused by gravitational acceleration, g.
Therefore, F = Mg. If the body is lifted a height h (Fig. 8) and friction is negligible, w = Mgh. For problems in which h is mich less than the radius from the center of the earth ($h << r$), g may be considered a constant.

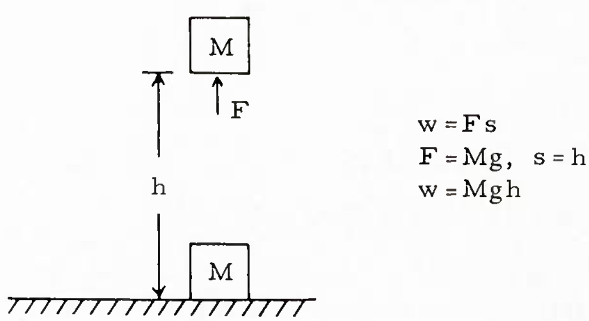


Figure 8. Work performed in lifting.

If an object is pushed up a frictionless inclined plane, the work done is still Mgh (Fig. 9).

$$w = Fs$$
$$F = Mg \sin \theta$$
$$s = \frac{h}{\sin \theta}$$
$$w = Mgh$$

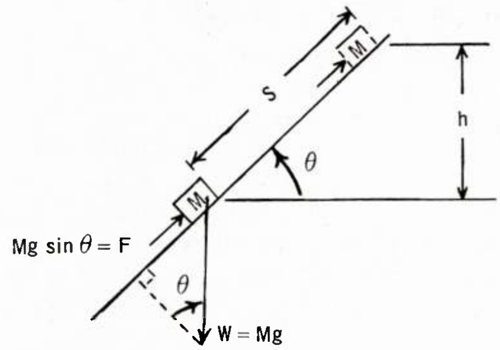


Figure 9. Work performed on a frictionless inclined plane.

For orbital mechanics problems, g varies and must be replaced by the value $\sqrt{g_1 g_2}$ where the subscripts indicate the beginning and final values of g. In such cases

$$w = M\sqrt{g_1 g_2}\,h$$

Another type of work is that work done against inertia. If, in moving from one point to another, the velocity of a body is changed, work is done. This work against inertia is computed in the following steps:

$$w = Fs$$
$$\text{but, } F = M\,a$$
$$\text{and } 2as = v_f^2 - v_o^2$$
$$s = \frac{v_f^2 - v_o^2}{2a}$$
$$\text{so, } w = Fs = M\,\frac{a\,(v_f^2 - v_o^2)}{2a} = \frac{M\,(v_f^2 - v_o^2)}{2} = \frac{Mv_f^2}{2} - \frac{Mv_o^2}{2}$$

The quantity $\frac{Mv^2}{2}$ is defined as kinetic energy (KE). Therefore, work done against inertia (if the altitude and the mass remain the same) is equal to the change in kinetic energy. Energy is defined as the ability to do work, and it is obvious that a moving body has the ability to do work (for example, a moving hammer's ability to drive a nail). A body is also able to do work because of its position or altitude; this is known as potential energy (PE). Units used to measure energy are similar to those used to measure work in that both are scalar rather than vector quantities.

The sum of the kinetic and the potential energy of a body is its *total mechanical energy.*

Newton's Law of Universal Gravitation

Newton published his *Principia* in 1687 and included in it the law of universal gravitation, which he had been considering for about twenty years. This law was based on observations made by Newton. Later work showed that it was only an approximation, but an extremely good approximation. The law states: *Every particle in the universe attracts every other particle with a force that is proportional to the product of the masses and inversely proportional to the square of the distance between the particles.* A constant of proportionality, G, termed the Universal Gravitational Constant, was introduced, and the law was written in this manner:

$$F = \frac{Gm_1 m_2}{r^2}$$

The value of G, the Universal Gravitational Constant, was first determined by Cavandish in a classical experiment using a torsion balance. The value of G is quite small ($G = 6.6695 \times 10^{-8}$ cgs units). In most problems the mass of one of the bodies is quite large. It is convenient, therefore, to combine G and the large mass, m_1, into a new constant, μ (mu), which is defined as the gravitational parameter. This parameter has different values depending upon the value of the large mass, m_1. If m_1 refers to the earth, the gravitational parameter, μ, will apply to all earth satellite problems. However, if the problem concerns satellites of the sun or other large bodies, μ will have a different value based on the mass of that body.

If we now simplify the law of gravitational attraction by combining G and m_1 and by adjusting the results for the English engineering unit system, we obtain the following:

$$G m_1 = \mu \frac{ft^3}{sec^2}$$

$$F = \frac{\mu}{r^2} m \text{ (Where F is lb force and m is slugs)}$$

If this expression is equated to the expression of Newton's Second Law of Motion, as it applies in a gravitational field, we see that:

$$F = mg = \frac{\mu}{r^2} m$$

and after dividing by the unit mass, m, we obtain:

$$g = \frac{\mu}{r^2}$$

Thus, the value of g varies inversely as the square of the distance from the center of the attracting body.

For problems involving earth satellites, the following two constants are necessary for a proper solution:

$$G m_{earth} = \mu_{earth} = 14.08 \times 10^{15} \frac{ft^3}{sec^2}$$

$$r_e \text{ (radius of earth)} = 20.9 \times 10^6 \text{ ft}$$

The formulas must be used with proper concern for the units involved, and the value given for μ applies only to bodies attracted to the earth.

Before applying Newton's law of universal gravitation to the solution of problems, it would be well to consider the possible paths that a body in unpowered flight must follow through space.

CONIC SECTIONS

The conic sections were studied by the Greek mathematicians, and a body of knowledge has accumulated concerning them. They have assumed new significance in the field of astronautics because *any free-flight trajectory can be*

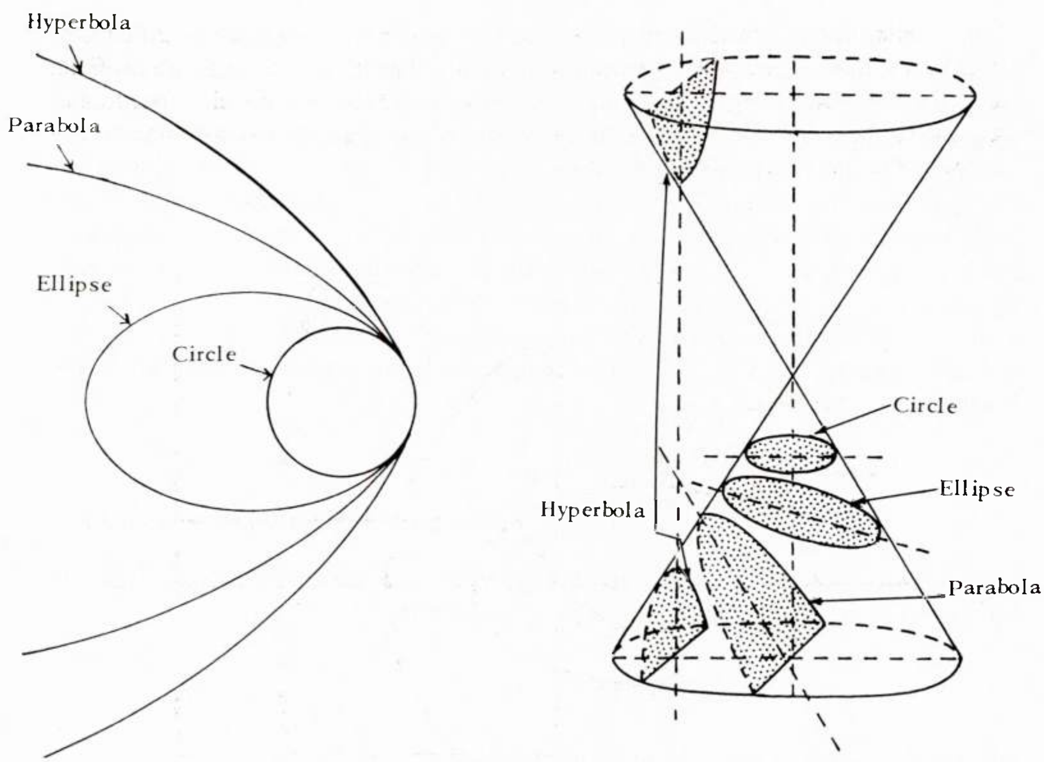


Figure 10. Conic sections.

represented by a conic section. The study of conic sections, or conics, is part of analytic geometry, a branch of mathematics that brings together concepts from algebra, geometry, and trigonometry.

A conic section is a curve formed when a plane cuts through a right circular cone at any point except at the vertex, or center. If the plane cuts both sides of one nappe of the cone, the section is an *ellipse* (Fig. 10). The *circle* is a special case of the ellipse occurring when the plane cuts the cone perpendicularly to the axis. If the plane cuts the cone in such a way that it is parallel to one of the

sides of the cone, the section is called a *parabola.* If the plane cuts both nappes of the cone, the section is a *hyperbola* which has two branches.

In one mathematical sense, all conic sections can be defined in terms of eccentricity (ϵ). The numerical value of ϵ is an indication of the relative shape of the conic (rotund or slender) and also an indication of the identity of the conic.

If the eccentricity is zero, the conic is a circle; if the eccentricity is greater than zero but less than one, the conic is an ellipse; if the eccentricity is equal to one, the conic is a parabola; and if the eccentricity is greater than one, the conic is a hyperbola.

Conic Sections and the Coordinate Systems

In locating orbits or trajectories in space, it is possible to make use of either rectangular (sometimes called Cartesian) or polar coordinates. In dealing with artificial satellites, it is often more convenient to use polar rather than rectangular coordinates because the center of the earth can be used both as the origin of the coordinates and as one of the foci of the ellipse.

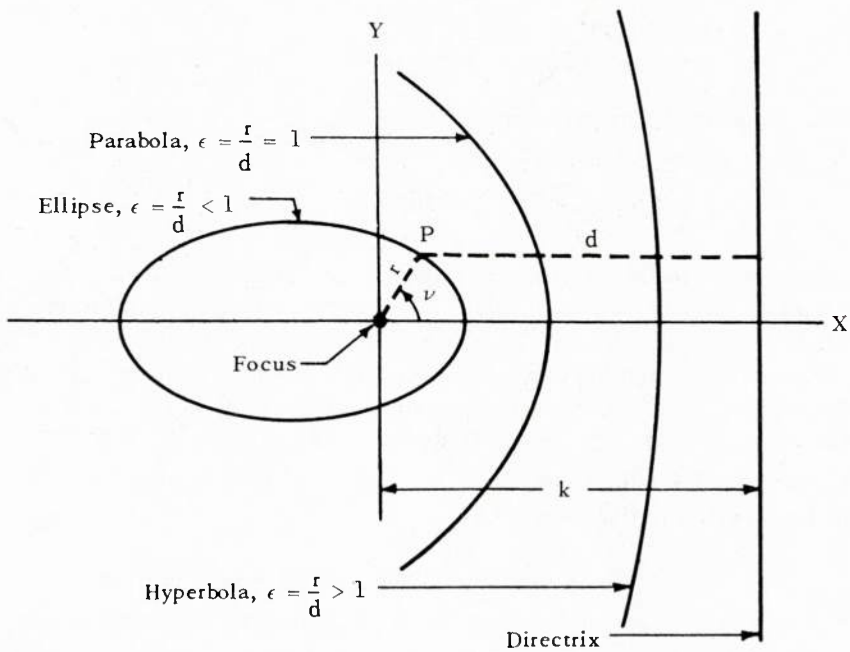


Figure 11. Rectangular and polar coordinates superimposed on the conic sections.

If rectangular and polar coordinates are superimposed upon a set of conics as shown in Figure 11, equations of the curves can be derived.

The formula for the eccentricity of a conic is $\epsilon = \frac{r}{d}$. This ratio is constant for a specific curve.
In Cartesian coordinates:

$$\epsilon = \frac{r}{d} = \frac{\sqrt{x^2 + y^2}}{k - x}$$

$$\sqrt{x^2 + y^2} = \epsilon(k - x)$$

Squaring both sides gives:

$$x^2 + y^2 = \epsilon^2(k - x)^2$$

To convert the rectangular coordinates to polar coordinates, substitute as follows:

$$x = r \cos \nu \text{ (ν is the lower case Greek nu)}$$

$$y = r \sin \nu$$

$$\epsilon = \frac{r}{d} = \frac{r}{k - r \cos \nu}$$

$$k\epsilon - r\epsilon \cos \nu = r$$

$$r + r\epsilon \cos \nu = k\epsilon$$

$$r = \frac{k\epsilon}{1 + \epsilon \cos \nu}$$

This result is the general equation for *all* conics.

Ellipse

The ellipse is the curve traced by a point (P) moving in a plane such that the sum of its distances from two fixed points (foci) is constant. In the ellipse in Figure 12, the following are shown: the foci (F and F′); c, distance from origin to either focus; a, distance from origin to either vertex (semimajor axis); 2a, major axis; b, distance from origin to intercept on y-axis (semiminor axis); 2b, minor axis; and $r + r'$, distances from any point (P) on the ellipse to the respective foci (F and F′).

A number of relationships which are very useful in astronautics are derived from the geometry of the ellipse:

$$r + r' = 2a \text{ (at any point on the ellipse)}$$

$$a^2 = b^2 + c^2 \text{ or } a = \sqrt{b^2 + c^2}$$

$$b = \sqrt{a^2 - c^2}$$

$$c = \sqrt{a^2 - b^2}$$

The eccentricity of the ellipse $(\epsilon) = \frac{c}{a}$. A chord through either focus perpendicular to the major axis is called the *latus rectum* and its length $= \frac{2b^2}{a}$.

These relationships can be used to determine the parameters of an elliptical orbit of a satellite when only the radius of perigee and the radius of apogee are known. These parameters are important because, as is shown later, they

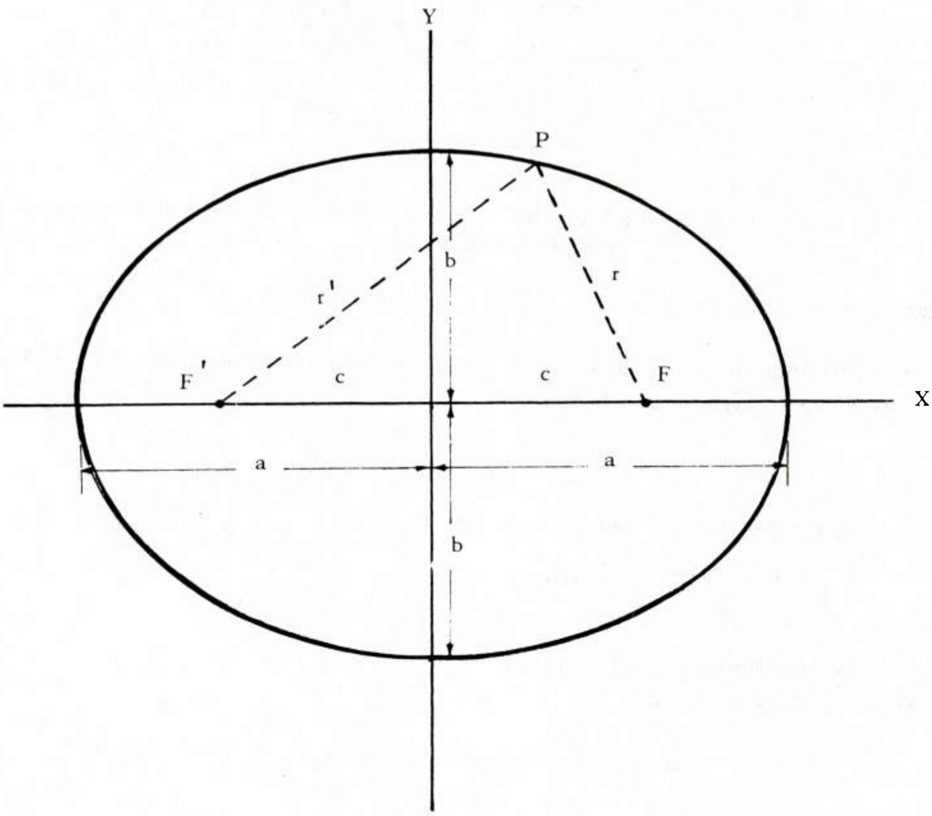


Figure 12. Ellipse with center at origin of rectangular coordinate system.

are related to the total mechanical energy and total angular momentum of the satellite. Thereby they offer a means of determining these values through the simple arithmetic of an ellipse rather than the vector calculus of celestial mechanics.

Sample problem: A satellite in a transfer orbit has a perigee at 300 NM above the surface of the earth and an apogee at 19,360 NM. Find a, b, c, and ϵ for the ellipse traced out by this satellite.

Solution:

Since the center of the earth is one focus of the ellipse, first convert the apogee and perigee to radii by adding the radius of the earth (3440 NM):

$$\begin{aligned}\text{radius of perigee } r_p &= \text{altitude of perigee} + \text{radius of earth}\\ &= 300 + 3440 = 3740 \text{ NM.}\\ \text{radius of apogee } r_a &= \text{altitude of apogee} + \text{radius of earth}\\ &= 19{,}360 + 3440 = 22{,}800 \text{ NM.}\end{aligned}$$

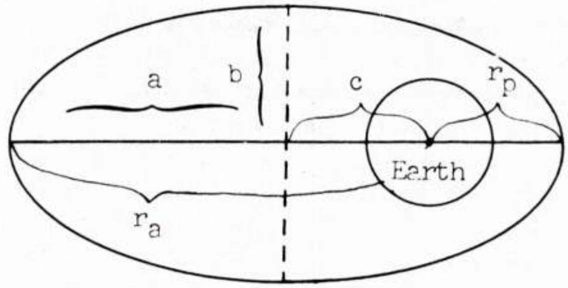


Figure 13. Orbit of an artificial satellite showing radius of perigee and radius of apogee (not to scale).

With this information, an exaggerated sketch of the ellipse can be made (Fig. 13). Compare this with Figure 12 to obtain:

$$r_p + r_a = \text{major axis} = 2a$$

then $\quad 2a = 3740 + 22{,}800 = 26{,}540 \text{ NM}$

or $\quad a = \dfrac{26{,}540}{2} = 13{,}270 \text{ NM}$

Also from comparing Figures 12 and 13:

$$\begin{aligned} c &= a - r_p \\ &= 13{,}270 - 3740 = 9530 \text{ NM} \end{aligned}$$

Since a and c are known, find b from the relationship given:

$$\begin{aligned} b &= \sqrt{a^2 - c^2} \\ \text{or } b &= \sqrt{(1.327 \times 10^4)^2 - (.953 \times 10^4)^2} \\ b &= \sqrt{(1.761 \times 10^8) - (.908 \times 10^8)} = \sqrt{.853 \times 10^8} \\ b &= .923 \times 10^4 = 9230 \text{ NM} \end{aligned}$$

According to the formula given for eccentricity:

$$\epsilon = \frac{c}{a}$$

$$\epsilon = \frac{9530}{13270} = .718$$

The ellipse is a conic section with eccentricity less than 1 ($\epsilon < 1$).

CIRCLE. The circle is a special case of an ellipse in which the foci have merged at the center; thus $\epsilon = 0$. The ellipse relationships can be used for a circle.

ENERGY AND MOMENTUM

Once the basic geometry of a trajectory or orbit is understood, the next subject for investigation is the physics of energy and momentum. From concepts of linear and angular motion, concepts of linear and angular momenta logically follow. Once the formulas for computing the specific angular momentum and the specific mechanical energy of a body in orbit are delineated, then it is possible to solve for unknown quantities, such as the altitude of the body above the surface of the earth or the velocity at any point on the orbit. Any body in space following a free-flight path—whether it is a missile, a satellite, or a natural body—is governed by the laws of the conservation of specific mechanical energy and specific angular momentum. Once the value of either of these items is known at any point along a free-flight trajectory or orbit, then its value is known at all other points, since the value does not change unless the body is acted upon by some outside force.

Mechanical Energy

The law of Conservation of Energy states that *energy can neither be created nor destroyed but only converted from one form to another.* This law can be applied to orbital mechanics and restated in this way: *The total mechanical energy of an object in free motion is constant, provided that no external work is done on or by the system.* During reentry, work is done by the system and some of the mechanical energy is converted to heat. Similarly, during launch, work is done on the system as the propulsion units give up chemical energy. In this chapter, only the free-flight portion of the trajectory is considered, and it is assumed that there is no thrust and no drag.

In order to establish a common understanding about changes in the amount of energy, it is necessary to agree upon a zero reference point for energy. Potential energy, or energy due to position, can be, and often is, measured from sea level. In working with earth-orbiting systems, however, the convention is to consider a body as having zero potential energy if it is at an infinite distance from the earth and as having zero kinetic energy if it is absolutely at rest with respect to the center of the earth. Under these circumstances, the total mechanical energy (PE + KE) is also equal to zero. If the total mechanical energy is positive—that is, larger than zero—the body has enough energy to escape from the earth. If the total mechanical energy is negative—that is, less than zero—the body does not have enough energy to escape from the earth, and it must be either in orbit or on a ballistic trajectory.

The formula for PE, with the reference system as stated above, is PE $= \frac{-m\mu}{r}$. Instead of using PE, a specific PE (PE per unit mass) can be used if both sides are divided by m; for example

$$\frac{PE}{m} = \frac{-\mu}{r}$$

If a body is at infinity, it has a specific PE equal to $-\frac{\mu}{r} = \frac{\mu}{\infty} = 0.$

A similar case can be presented for kinetic energy. A body with some velocity relative to the center of the earth has kinetic energy defined by:

$$KE = \frac{mv^2}{2}$$

Again, the specific kinetic energy (kinetic energy per unit mass) can be defined as:

$$\text{Specific KE} = \frac{KE}{m} = \frac{v^2}{2}$$

In general, a body in free motion in space has a particular amount of mechanical energy, and this amount is constant because of the conservation of mechanical energy.

$$\text{Total Mechanical Energy} = KE + PE$$

A more useful expression is obtained if we define Specific Mechanical Energy, E, or the Total Mechanical Energy per unit mass. Thus, we can write:

$$E = \frac{\text{Total Mechanical Energy}}{m}$$

$$E = \frac{KE}{m} + \frac{PE}{m}$$

$$E = \frac{v^2}{2} - \frac{\mu}{r}$$

Specific Mechanical Energy, E, is also conserved in unpowered flight in space. The units of E are $\frac{ft^2}{sec^2}$. Since the mass term does not appear directly in the equation, E represents the specific mechanical energy of a body in general.

If the solution to the Specific Mechanical Energy equation yields a negative value for E, the body is on an elliptical or circular path (nonescape path). If E is exactly equal to zero, the path is parabolic; this is the minimum energy escape path. If E is positive, the path is hyperbolic, and the body will also escape from the earth's gravitational field.

Although the value of E, once determined, remains constant in free flight, there is a continuous change in the values of specific kinetic energy and specific potential energy. High velocities nearer the surface of the earth, representing high specific kinetic energies, are exchanged for greater specific potential energies as distance from the center of the earth increases. In general, velocity is traded for altitude; kinetic energy is traded for potential energy. The sum remains constant.

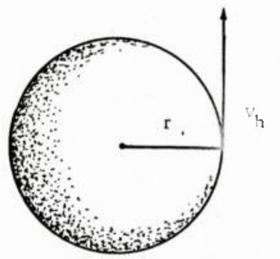
Linear and Angular Momentum

When a body is in motion, it has momentum. Momentum is the property a body possesses because of its mass and its velocity. In linear motion, momentum is expressed as mv and has the units, $\frac{\text{foot-slug}}{\text{sec}}$.

When a rigid body, such as a flywheel, rotates about a center, it has angular momentum. Once the flywheel is in motion, its angular momentum would remain constant if it were not acted upon by forces such as friction and air resistance. Similarly, a gyroscope would rotate indefinitely in the absence of friction and air resistance. Thus, ignoring such losses, angular momentum will remain constant. In space, it can be assumed that such forces are negligible and that angular momentum is conserved. This is another tool to use in analyzing orbital systems.

Angular momentum is the product of moment of inertia, I, and the angular velocity, ω. Moment of inertia of a body of mass, m, rotating about a center at a distance, r, can be expressed as m r^2. The angular momentum is then equal to m r^2 ω.

For convenience in calculations, the term Specific Angular Momentum, H, is defined as the angular momentum per unit mass. Remembering that the magni-



$$\text{Angular Momentum} = m\,r^2\,\omega$$

$$\omega = \frac{v_{horizontal}}{r}$$

$$H = \frac{\text{Ang. Momentum}}{m}$$

$$\therefore H = v_h r \text{ (Circular Motion)}$$

Figure 14. Specific angular momentum of a circular orbit.

tude of the instantaneous velocity vector of a body rotating in constant circular motion about a center with radius r is equal to ωr and that the vector is perpendicular to the radius, the expression for specific angular momentum of a circular orbit can be simplified as shown in Figure 14.

The general application of specific angular momentum to all orbits requires that the component of velocity perpendicular to the radius vector be used. This velocity component is defined as

$$v_h = v \cos \phi$$

where ϕ is the angle the velocity vector makes with the local horizontal, a line perpendicular to the radius. In an elliptical orbit, the geometry is as shown in Figure 15. The body in orbit has a total velocity v which is always tangent to the flight path.

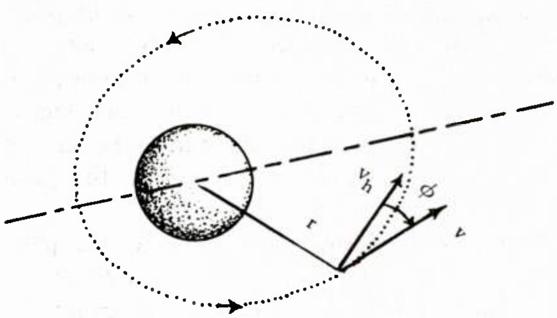
The formula, $H = v\,r \cos \phi$, defines the specific angular momentum for all orbital cases. The angle ϕ *is the flight path angle* and is the angle between the local horizontal and the total velocity vector. It should be noted that the angle ϕ is equal to zero for circular orbits. Further, in elliptical orbits, ϕ is zero at the points of apogee and perigee.

The two important formulas that have been presented in this section are those for E and H. These formulas permit a trajectory or an orbit to be completely defined from certain basic data:

$$E = \frac{v^2}{2} - \frac{\mu}{r}, \text{ when units of E are } \frac{ft^2}{sec^2}$$

$$H = v\,r\cos\phi, \text{ when units of H are } \frac{ft^2}{sec}$$

If v, r, and ϕ are known for a given trajectory (or orbit) at a given position, then E and H can be determined. In the absence of outside forces, E and H are constants; therefore v, r, and ϕ can be determined at any other position on the trajectory or orbit. Equations for the specific angular momentum and the specific mechanical energy can be used in practical application to the two-body problem and to the free-flight portion of the ballistic missile trajectory.



$H = v_h r$

but $v_h = v\cos\phi$

$\therefore H = v\,r\cos\phi$ (General orbit)

Figure 15. Specific angular momentum.

THE TWO-BODY PROBLEM

It is implicit in Newton's law of universal gravitation that every mass unit in the universe attracts, and is attracted by, every other mass unit in the universe. Clearly, small masses at large distances are infinitesimally attracted to each other. It is neither feasible nor necessary to consider mutual attractions of a large number of bodies in many astronautics problems. The most frequent problems of astronautics involve only two interacting bodies: a missile payload, or satellite, and the earth. In these instances, the sun and moon effects are negligible except in the case of a space probe, which will be noticeably affected by the moon, if it passes close to the moon, and which will be controlled by the sun, if it escapes from the earth's gravitational field.

Military officers concerned with operational matters are primarily interested in launching a missile from one point on the earth's surface to strike another point on the earth's surface and in launching earth satellites. In these problems, the path followed by the payload is adequately described by considering only two bodies, the earth and the payload. The problem of two bodies is termed the two-body problem; its solution dates back to Newton.

It is indeed fortunate that the solution of the two-body trajectory is simple and straightforward. A general solution to a trajectory involving more than two bodies does not exist. Special solutions for these more complex trajectories usually require machine calculation.

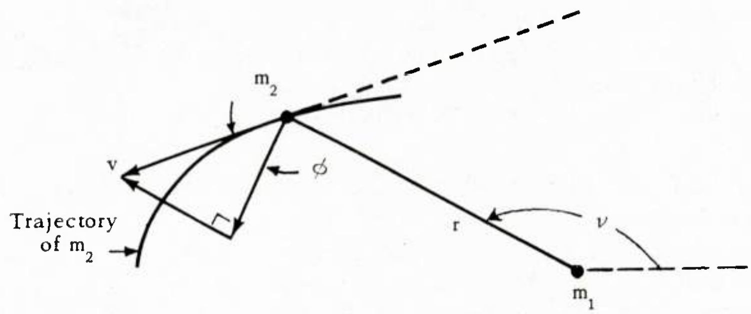


Figure 16. Trajectory relationships.

The two-body problem is described graphically in Figure 16. A small body, m_2, has a velocity, v, at a distance r from the origin chosen as the center of mass of a very massive body, m_1. The problem is to establish the path followed by body, m_2, or to define its trajectory. This is a typical problem in mechanics—given the present conditions of a body, what will these conditions be at any time, t, later? First, we shall find r as a function of ν, where ν is the polar angle measured from a reference axis to the radius vector.

At the outset, it should be apparent that the entire trajectory will take place in the plane defined by the velocity vector and the point origin. There are no forces causing the body, m_2, to move out of this plane; otherwise, the conditions are not those of a two-body free-flight problem.

In the earlier outline of the laws of conservation of energy and momentum, the following conditions were established:

$$\frac{v^2}{2} - \frac{\mu}{r} = E = \text{a constant} \tag{1}$$

$$H = vr \cos \phi = \text{a constant} \tag{2}$$

Equations (1) and (2) can be combined and, with the aid of the calculus, the following equation can be derived:

$$r = \frac{H^2/\mu}{1 + \sqrt{1 + \frac{2EH^2}{\mu^2}} \cos \nu} \tag{3}$$

Equation (3) is the equation of a two-body trajectory in polar coordinates.

Earlier, the following equation was given as the equation of any conic section in polar coordinates, the origin located at a focus:

$$r = \frac{k\epsilon}{1 + \epsilon \cos \nu} \tag{4}$$

Equations (3) and (4) are of the same form; hence, equation (3) is also the equation of any conic section (origin at a focus) in terms of the physical constants, E and H, and the two-body trajectories are then conic sections. This conclusion substantiates Kepler's first law. In fact, Kepler's first law is a special case because an ellipse is just one form of conic section.

Since equations (3) and (4) are of the same form, it is possible to equate like terms, which will lead to relationships between the physical constants, E, H, and μ, and the geometrical constants, ϵ, a, b, and c. Thus:

$$k\epsilon = \frac{H^2}{\mu} \tag{5}$$

and

$$\epsilon = \sqrt{1 + \frac{2EH^2}{\mu^2}} \tag{6}$$

Physical Interpretation of the Two-Body Trajectory Equation

Analysis of the two-body trajectory equation will give an understanding of the physical reaction of a vehicle (small body) under the influence of a planet (large body).

If $E < 0$, the trajectory is an ellipse. What is the condition that E be less than zero? It is simply that the kinetic energy of the small mass, m_2, because of its relatively low velocity, is less than the magnitude of its potential energy. Therefore, the body cannot possibly go all the way to infinity; that is, it cannot go to a point where it is no longer attracted by the larger body—where the potential energy is zero. The smaller mass cannot escape. It must remain "captured" by the force field of the larger body. Therefore, it will be turned back toward the larger body, or, more in keeping with the idea of potential energy, it will always be "falling back" toward the more massive body. When this particular balance of energy exists, the trajectory is elliptical with one focus coincident with the center of mass of the larger body. In the actual physical case, the larger body will have a finite size; that is, it will not be a point mass, and this ellipse may intersect the surface of the larger body as it does in the case of a ballistic missile. If the velocity is sufficiently high, and its direction proper, the ellipse may completely encircle the central body, the condition of a satellite.

If $E = 0$, the kinetic energy exactly equals the magnitude of the potential energy, and the small mass, m_2, has just enough energy to travel to infinity, away from the influence of the central body, and come to rest there. The small body will follow a parabolic path to infinity. The velocity which is associated with this very special energy level is also very special and is commonly called the "escape velocity."

Escape velocity can be calculated by setting $E = 0$ in the mechanical energy equation (1) as follows:

$$\frac{v^2_{esc}}{2} - \frac{\mu}{r} = E = 0$$

$$v^2_{esc} = 2\left(\frac{\mu}{r}\right)$$

$$v_{esc} = \sqrt{\frac{2\mu}{r}} \tag{7}$$

Thus, it can be seen that escape velocity decreases with distance from the center of the earth. At the earth's surface,

$$v_{esc} = \sqrt{\frac{2\mu}{r_e}} = \left[\frac{(2)\,(14.08)\,(10^{15}\,\frac{ft^3}{sec^2})}{(20.9)(10^6\,ft)}\right]^{\frac{1}{2}}$$

$$v_{esc} = 36{,}700 \text{ ft/sec.}$$

If the velocity of the small mass exceeds escape velocity, which will be the case if $E > O$, it will follow a hyperbolic trajectory to infinity. In practice infinity is a large distance at which the earth's attractive force is insignificant, and there the mass will have some residual velocity. In a mathematical sense, the body would still have velocity at infinity. In a physical sense, it would have velocity relative to the earth at any large distance from the earth.

Considering the sounding rocket, only the straight-line, degenerate conic is a possible trajectory. But, again, the value of E will determine whether escape is possible; that is, if $E < O$, the straight-line trajectory cannot extend to infinity. If $E = O$ or $E > O$, the straight line will extend to infinity.

Example Problem: The first U.S. "moon shot," the Pioneer I, attained a height of approximately 61,410 NM above the earth's surface. Assuming that the Pioneer had been a sounding rocket (a rocket fired vertically), and assuming a spherical, nonrotating earth without atmosphere, calculate the following:

a. E (total specific energy)

b. Impact velocity (earth's surface)

Solution: Given

(a) At apogee (greatest distance from earth):

Altitude (above earth's surface) = 61,410 NM

Earth radius = 3440 NM

Velocity = O (Only for a sounding rocket)

r = altitude + earth's radius

r = (61,410 + 3440) NM = 64,850 NM

$$E = \frac{v^2}{2} - \frac{\mu}{r} = 0 - \frac{(14.08)\,(10^{15})\,\frac{ft^3}{sec^2}}{(64{,}850)\text{ NM }(6080)\,\frac{ft}{NM}}$$

$$E = -\,3.57 \times 10^7 \text{ ft}^2/\text{sec}^2$$

Answer

(b) Since the specific energy is constant,

At the earth's surface:

$r = 3440$ NM

$E = -\,3.57 \times 10^7$ ft^2/sec^2

$$E = \frac{v^2}{2} - \frac{\mu}{r}$$

$$-3.57 \times 10^7 \frac{ft^2}{sec^2} = \frac{v^2}{2} - \frac{(14.08)\,(10^{15}\frac{ft^3}{sec^2})}{(3440)\,NM\,(6080)\,\frac{ft}{NM}}$$

$$v^2_{(impact)} = 2\,[(-3.57 \times 10^7) + (67.4 \times 10^7)] = 2(63.8 \times 10^7) ft^2/sec^2$$

$$v_{(impact)} = 35{,}700\ ft/sec \qquad \text{Answer}$$

This is also the approximate burnout velocity of the vehicle. As the surface escape velocity is 36,700 ft/sec, it is clear that Pioneer I did not attain escape velocity, and so it returned to earth.

Elliptical Trajectory Parameters

While parabolic and hyperbolic trajectories, especially the latter, are of interest in problems of interplanetary travel, elliptical trajectories comprise the ballistic missile and satellite cases, which are of current military interest. It is important, then, to relate the dimensions of an ellipse (a, b, and c) to the physical constants (E, H, and μ) as was previously done for ϵ.

The relationship, $r_a + r_p = 2a$, was presented earlier. If this equation is applied to point P in Figure 17,

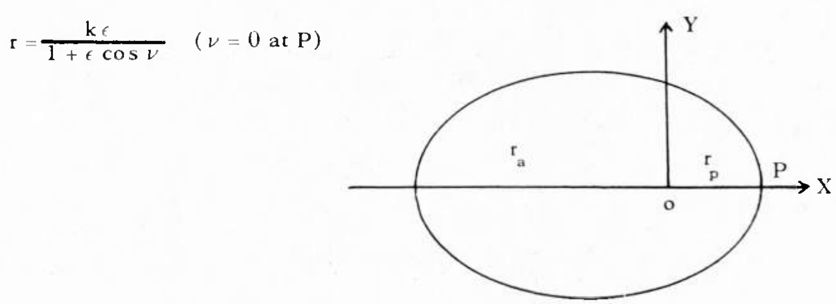


Figure 17. Ellipse.

$$r = \frac{k\epsilon}{1 + \epsilon} \text{ and } r_p = a - c. \text{ Then,}$$

$$\frac{k\epsilon}{1 + \epsilon} = a - c \tag{8}$$

But $\epsilon = \frac{c}{a}$; therefore, substituting $c = \epsilon a$ into (8),

$$\frac{k\epsilon}{1 + \epsilon} = a - \epsilon a = a(1 - \epsilon)$$

$$k\epsilon = a(1 - \epsilon^2) \tag{9}$$

But from equations (5) and (6),

$$k\epsilon = H^2/\mu \text{ and } \epsilon = \sqrt{1 + \frac{2EH^2}{\mu^2}}$$

Substituting these relationships into equation (9),

$$\frac{H^2}{\mu} = a\left(1 - 1 - \frac{2EH^2}{\mu^2}\right)$$

$$\frac{H^2}{\mu} = -\frac{2EH^2 a}{\mu^2}$$

$$-1 = \frac{2Ea}{\mu}$$

$$a = -\frac{\mu}{2E} \quad \text{(EXTREMELY USEFUL)} \tag{10}$$

$$\text{Also } a = -\frac{\mu}{2\left(\frac{v^2}{2} - \frac{\mu}{r}\right)} \tag{11}$$

From the following equation:

$$\frac{b^2}{a^2} = 1 - \epsilon^2$$

$$\frac{b^2}{a} = a(1 - \epsilon^2)$$

But from equation (9),

$$a(1 - \epsilon^2) = k\epsilon = \frac{H^2}{\mu}$$

Therefore,

$$\frac{b^2}{a} = \frac{H^2}{\mu} \tag{12}$$

Equations (10) and (12) are extremely important relationships; an understanding of them is essential to material that follows on ballistic missiles and satellites. If injection conditions of speed and radius are fixed, it is clear that a, the semimajor axis of the elliptical trajectory, becomes fixed, regardless of the value of the flight path angle at burnout. Equation (10) points out there is a direct relationship between the size of an orbit and the energy level of the orbiting object. Equation (12) points out that for a given energy level there is a direct relationship between the

length of the semi latus rectum of an elliptical trajectory (a shape parameter) and the specific angular momentum of the orbiting object. This implies that the size and shape of an elliptical trajectory are determined by the E and H.

Two-Body Trajectory Definitions and Geometry

The general equation of two-body trajectories has now been introduced. Before proceeding to problem applications it would be well to consider in detail some commonly used terms and symbols.

First, refer to Figure 18. In general, the point P is called the periapsis and P′ the apoapsis. If the earth is at point O, the ellipse would then represent the trajectory of an earth satellite; P is then termed perigee and P′ apogee. If the sun is at point O, the ellipse would represent a planetary orbit; P is then called perihelion and P′ aphelion.

In order to explain the use of the angle ν, the geometry of satellites will be discussed briefly. Figure 18 depicts a planetary orbit (not to scale).

In astronomy and celestial mechanics it is standard practice to measure a body's position from perihelion point P. There are several reasons for using perihelion, including the fact that perihelion of any body in the solar system except Mercury and Venus is closer to the earth's orbit than is the body's aphelion. In fact, for a highly eccentric orbit such as a comet's, the body would not be visible at aphelion. In order to conform to accepted practice, then, the angle nu (ν), measured from periapsis, has been introduced. This angle, which is called the *true anomaly,* is of considerable importance in time-of-flight calculations.

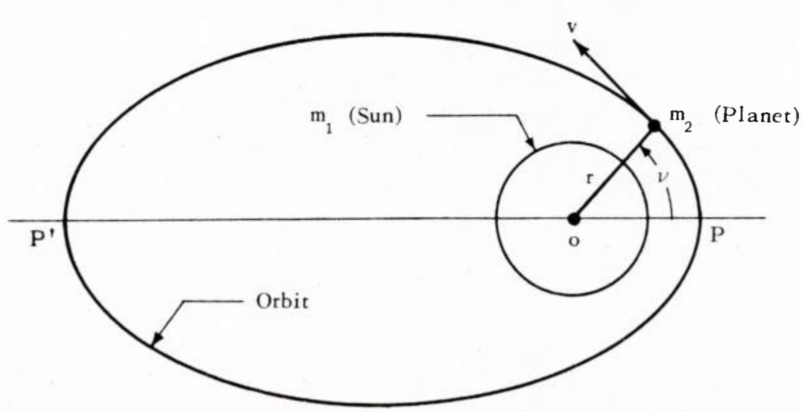


Figure 18. Sun-centered orbit

The *true anomaly,* however, does not lend itself well to ballistic missile problems, as can be seen from the simplified ballistic missile geometry in Figure 19.

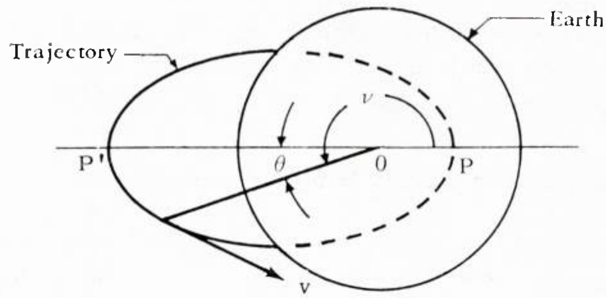


Figure 19. Ballistic missile trajectory.

In a ballistic missile trajectory, perigee is entirely fictitious. The missile obviously never traverses the dashed portion of the trajectory. The solid portion of the trajectory is all that is of real interest, and this portion is in the second and third quadrants of the angle ν. It is convenient then to define an angle θ, measured counterclockwise from apogee (apoapsis, in general), such that

$$\nu = \theta + \pi. \tag{13}$$

θ will normally have values in the first and fourth quadrants. From (13), it is clear that derivatives of ν and θ will be interchangeable. The equation of a conic section in terms of θ can be found by substituting (13) into (4),

$$r = \frac{k\epsilon}{1 + \epsilon \cos \nu} = \frac{k\epsilon}{1 + \epsilon \cos (\theta + \pi)}$$

$$r = \frac{k\epsilon}{1 - \epsilon \cos \theta}.$$

With this understanding of the relationship between ν and θ, it will be convenient to use ν when working with satellite and space trajectories and θ when working with ballistic missiles (see App. C).

EARTH SATELLITES

During their free flight trajectory, satellites and ballistic missiles follow paths described by the two-body equation. For a satellite to achieve orbit, enough energy must be added to the vehicle so that the ellipse does not intersect the surface of the earth. However, not enough energy is added to allow the vehicle to escape. Therefore, the ellipse and the circle are the paths of primary interest.

The orientation, shape, and size of orbits are important to the accomplishment of prescribed missions. Therefore, eccentricity (ϵ), major axis (2a), minor axis (2b), and distance between the foci (2c) are of interest. It is necessary to know the relationships of these geometric values to the orbital parameters in order to make an analysis of an orbit. For example, it is helpful to remember that:

r_p (radius at perigee) $= a - c$

r_a (radius at apogee) $= a + c$

$r_p + r_a = 2a$

$$\epsilon = \frac{c}{a}$$

$a^2 = b^2 + c^2$

Specific mechanical energy, E, and specific angular momentum, H, are of primary concern when elliptical and circular orbits are discussed. If there are no outside forces acting on a vehicle in an orbit, the specific mechanical energy and the specific angular momentum will have constant values, regardless of position in the orbit. This means that if E and H are known at one point in the orbit, they are then known at each and every other point in the orbit. At a given position if radius r, speed v, and flight path angle ϕ are known, E and H can be determined from:

$$E = \frac{v^2}{2} - \frac{\mu}{r} = -\frac{\mu}{2a}$$

$$H = vr \cos \phi$$

If the values of E and H are known for a particular orbit, and the speed and flight path angle at a certain point in the orbit are to be determined, the energy equation can be solved for v, and then the angular momentum equation can be solved for ϕ.

The equations for the speed in circular and elliptical orbits are important. The equation for circular speed is:

$$v = \sqrt{\frac{\mu}{r}}$$

The equation for elliptical speed is:

$$v = \sqrt{\frac{2\mu}{r} - \frac{\mu}{a}}$$

Another equation that is important in the analysis of orbits is the equation for orbital period. For a circular orbit the distance around is the circumference of the circle which is $2\pi r$. Therefore, the period, which is equal to the distance around divided by the speed, is this:

$$P = \frac{2\pi r}{v}.$$

Now, substitute for v the speed in circular orbit:

$$v = \sqrt{\frac{\mu}{r}}$$

$$P = \frac{2\pi r}{\sqrt{\frac{\mu}{r}}}$$

Multiply numerator and denominator of the right hand side by $\sqrt{r}$:

$$P = \frac{2\pi r \sqrt{r}}{\sqrt{\frac{\mu}{r}} \sqrt{r}} = \frac{2\pi r^{3/2}}{\sqrt{\mu}}$$

Using the principle of Kepler's third law, replace r by the mean distance from the focus, which is equal to the semimajor axis a, and the equation becomes:

$$P = \frac{2\pi a^{3/2}}{\sqrt{\mu}}$$

Squaring both sides, $P^2 = \frac{4\pi^2 a^3}{\mu}$. Since $\frac{4\pi^2}{\mu}$ is a constant, P^2 is proportional to a^3, and for earth satellites $P^2 = \left(2.805 \times 10^{-15} \frac{sec^2}{ft^3}\right) (a)^3$. Or, $P = \left(5.30 \times 10^{-8} \frac{sec}{ft^{\frac{3}{2}}}\right) (a)^{\frac{3}{2}}$.

Problem: Initial data from Friendship 7 indicated that the booster burned out at a perigee altitude of 100 statute miles, speed of 25,700 ft/sec, and flight path angle of 0°. Determine the speed and height at apogee, and the period.

Given*: $h_p = 100\ SM = .5 \times 10^6\ ft$ $\qquad r_e = 20.9 \times 10^6\ ft$

$v_{bo} = 25{,}700\ ft/sec$

$\phi_{bo} = 0°$

Find: v_a, h_a, P

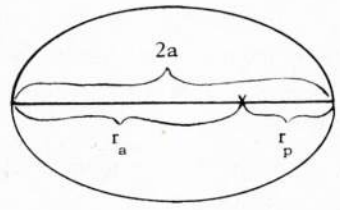


Figure 20. Orbit of Friendship 7 (not to scale).

* Even though $\phi_{bo} = 0°$, if burnout altitude were not given as perigee altitude, you would have to determine if this were perigee or apogee. To do this, you would compute the circular speed for the given burnout altitude and compare this with the actual speed. If the circular speed is greater than the actual, burnout was at apogee; if the circular speed was less than the actual speed, burnout was at perigee.

Solution:

$$E = \frac{v^2}{2} - \frac{\mu}{r} = \frac{(2.57 \times 10^4)^2}{2} - \frac{14.08 \times 10^{15}}{(20.9 + .5) \times 10^6}$$

$$= (3.31 \times 10^8) - (6.58 \times 10^8) = -3.27 \times 10^8$$

$$\text{but } E = -\frac{\mu}{2a};\ 2a = -\frac{\mu}{E} = \frac{-14.08 \times 10^{15}}{-3.27 \times 10^8}$$

$$= 43.1 \times 10^6$$

From Figure 20, $r_a + r_p = 2a$

$$\therefore r_a = 2a - r_p = (43.1 - 21.4) \times 10^6$$

$$= 21.7 \times 10^6$$

$$h_a = r_a - r_e = (21.7 - 20.9) \times 10^6 = .8 \times 10^6 \text{ ft}$$

$$= 151 \text{ sm} \qquad \text{Answer}$$

$$H_p = H_a$$

$$v_p r_p = v_a r_a \qquad v_a = \frac{v_p r_p}{r_a}$$

$$v_a = \frac{(2.57 \times 10^4)(21.4 \times 10^6)}{21.7 \times 10^6} = 25{,}400 \text{ ft/sec}$$

$$= 17{,}300 \text{ mph} \qquad \text{Answer}$$

$$P^2 = \frac{4\pi^2 a^3}{\mu} = (2.805 \times 10^{-15})(21.6 \times 10^6 \text{ ft})^3 \frac{\text{sec}^2}{\text{ft}^3}$$

$$= 28.1 \times 10^6 \text{ sec}^2$$

$$P = 5.30 \times 10^3 \text{ sec} = 88.3 \text{ min} \quad \text{Ans.}$$

It is interesting to compare the computed apogee and period results with the actual orbit (later data gave a higher accuracy for burnout conditions):

Item	*Actual figures*	*Computed figures*
V_{bo}	25,728 ft/sec	25,700 ft/sec
h_{bo}	97.695 SM	100 SM
h_a	158.85 SM	151 SM
P	88.483 min	88.3 min

Note that using three significant figures results in $h_p = .5 \times 10^6$ ft, about 94 SM. Such errors are common using slide rule accuracy, but this problem does illustrate the techniques used.

From the foregoing problem, it is evident that the principles and relatively simple algebraic exprcssions presented thus far are extremely important. They cnable one to analyze a trajectory or orbit rather completely—with a slide rule for academic or generalized discussion purposes, or with a digital computer for system design and operation. The discussion has been, however, confined to the two dimensional orbital plane. Before discussing some of the more interesting facets of orbital mechanics, it is necessary to properly locate a payload in three dimensions.

LOCATING BODIES IN SPACE

In one of the coordinate systems for space used by engineers and scientists, the origin is the center of the earth. This is a logical choice since the center of the earth is a focus for all earth orbits.

With the center of the coordinate systems established, a reference frame is required on which angular measurements can be made with respect to the center. The reference frame should be regular in shape, and it should be fixed in space. A sphere satisfies the requirement of a regular shape. The sphere of the earth would be a handy reference if it were fixed in space, but it rotates constantly.

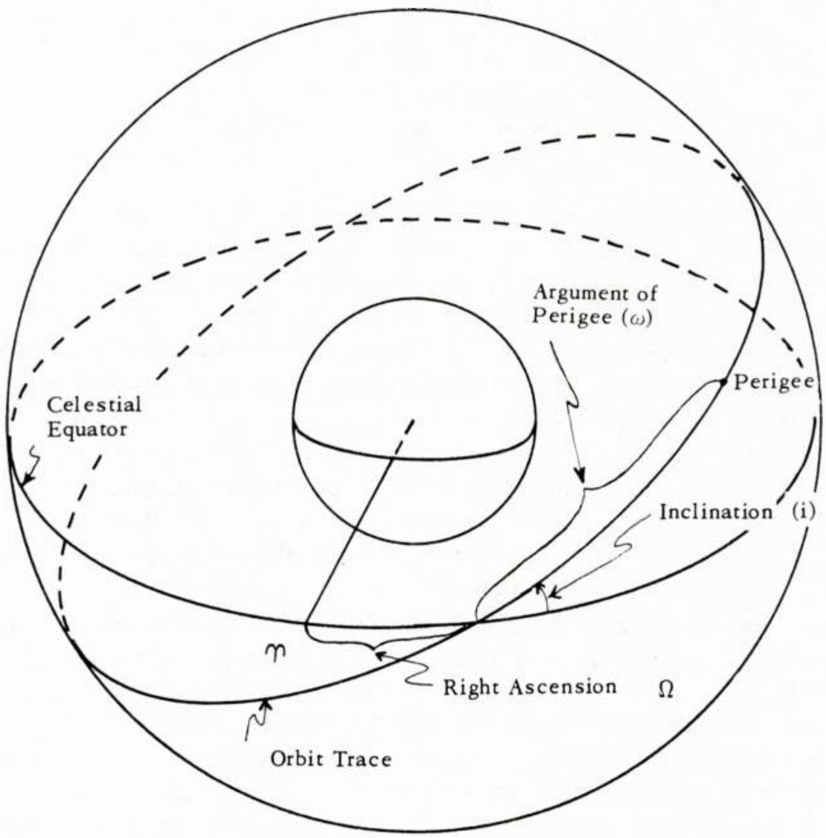


Figure 21. Celestial Sphere.

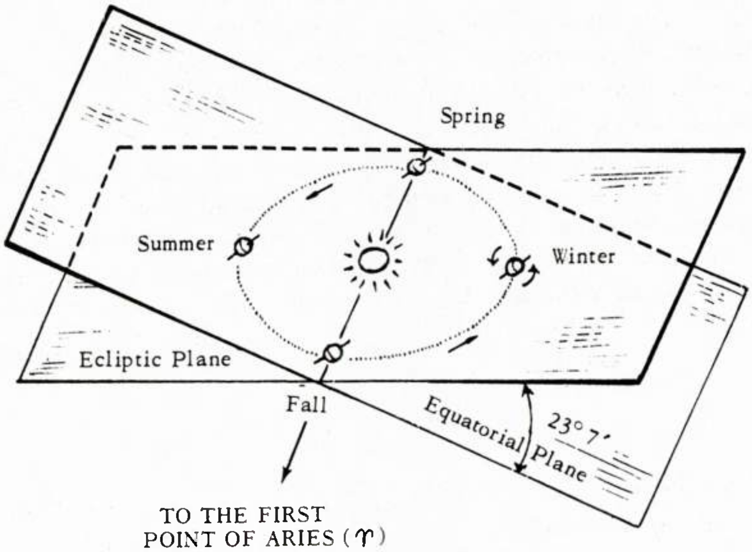


Figure 22. The vernal equinox.

Therefore, the celestial sphere is used to satisfy the requirement for a reference frame. This is a nonrotating sphere of infinite radius whose center coincides with the center of the earth and whose surface contains the projection of the celestial bodies as they appear in the sky (Fig. 21). The celestial equator is a projection of the earth's equator on the celestial sphere. The track of a satellite can be projected on the celestial sphere by extending the plane of the oribt to its intersection with the celestial sphere.

After the center of the system and the celestial equator have been defined, a reference is required as a starting point for position measurements. This point, determined at the instant winter changes into spring, is found by passing a line from the center of the earth through the center of the sun to the celestial equator, and is called the vernal equinox (Fig. 22).

After the references for the coordinate system have been established, the orbit itself must be located. The first item of importance is *right ascension* (Ω) of the ascending node, which is defined as the arc of the celestial equator measured eastward from the vernal equinox to the ascending node (Fig. 21). The ascending node is the point where the projection of the satellite path crosses the celestial equator from south to north. In other words, right ascension of the ascending node is the angle measured eastward from the first point of Aries to the point where the satellite crosses the equator from south to north.

The next item of importance is the angle the path of the orbit makes with the equator. This is the angle of inclination (i), which is defined as the angle that the plane of the orbit makes with the plane of the equator, measured counterclockwise from the equator at the ascending node. *Equatorial* orbits have $i = 0°$; *posigrade* orbits have $i = 0°$ to $90°$; *polar* orbits have $i = 90°$; and *retrograde* orbits have $i = 90°$ to $180°$.

To describe the orbit further, the perigee is located. The angular measurement from the ascending node to the perigee, measured along the path of the orbit in the direction of motion, is called the argument of perigee (ω).

If, in addition to the coordinates of the orbit, a time of either perigee or right ascension of the ascending node is known, along with the eccentricity and the major axis of the orbit, the exact position and velocity of the satellite can be determined at any time. Six quantities (right ascension, inclination, argument of perigee, eccentricity, major axis, and epoch time at either perigee or ascending node) form a convenient grouping of the minimum information necessary to describe the orbital path as well as the position of a satellite at any time. They constitute one set of orbital elements, known as the Breakwell Set of Keplerian Elements.

Orbital Plane

Another interesting facet of earth satellites concerns the orbital plane. There is a relationship between the launch site and the possible orbital planes. This restriction arises from the fact that the center of the earth must be a focus of the orbit and, therefore, must lie in the orbital plane.

The inclination of the orbital plane, i, to the equatorial plane is determined by the following formula: *cos i (inclination) = cos (latitude) sin (azimuth)* where the azimuth is the heading of the vehicle measured clockwise from true north.

As an example, a satellite launched from Cape Kennedy and injected at 30° N on a heading due east (Azimuth 90°) will lie in an orbital plane which is inclined 30° to the equatorial plane.

$$\begin{aligned}
\cos i &= (\cos \text{latitude})\ (\sin \text{azimuth}) \\
&= \cos 30^\circ \sin 90^\circ \\
\cos i &= \cos 30^\circ \\
i &= 30^\circ
\end{aligned}$$

It can be deduced from the above that the minimum orbital plane inclination for a direct (no dog leg or maneuvering) injection will be closely defined by the latitude of the launch site. All launch sites, thereby, will permit direct injections at inclination angles from that minimum (the approximate latitude of the launch site) to polar orbits (plus retrograde supplements), provided there were no geographic restrictions on launch azimuth, such as range safety limitations. For example, direct injections from Vandenberg AFB (35°N) would permit inclination angles from about 35° to 145°.

Once the inclination of the orbital plane is defined, the ground track can be discussed.

SATELLITE GROUND TRACKS

The orbits of all satellites lie in planes which pass through the center of a theoretically spherical earth. Each plane intersects the surface of the earth in a great circle (Fig. 23).

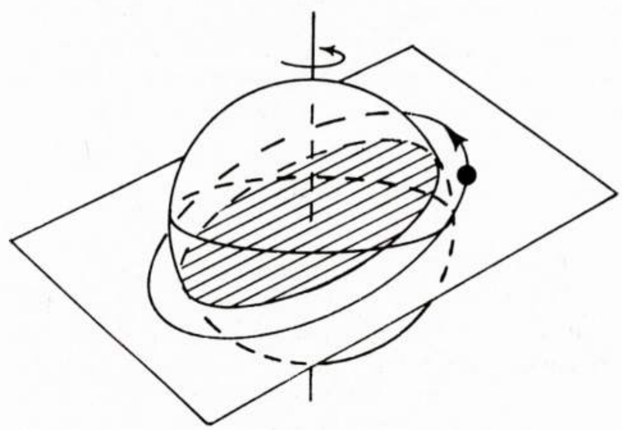


Figure 23. Satellite ground track geometry.

A satellite's ground track is formed by the intersection of the surface of the earth and a line between the center of the earth and the satellite. As the space vehicle moves in its orbit, this intersection traces out a path on the ground below.

There are five primary factors which affect the ground track of a satellite moving along a free flight trajectory. These are:

1. Injection point	
2. Inclination angle	(i)
3. Period	(P)
4. Eccentricity	(ϵ)
5. Argument of Perigee	(ω)

Of the above, the injection point simply determines the point on the surface from which the ground track begins, following orbital injection of the satellite. Inclination angle has been discussed in the previous section and will be treated below in further detail. Period, eccentricity, and argument of perigee each affect the ground track, but it is often difficult to isolate the effect of any one of the three. Therefore, only general remarks regarding the three factors will be made, rather than an intricate mathematical treatment.

If the study of satellite ground tracks is predicated upon a nonrotating earth, the track of a satellite in a circular orbit is easy to visualize. When the satellite's orbit is in the equatorial plane, the ground track coincides with the equator (Fig. 24).

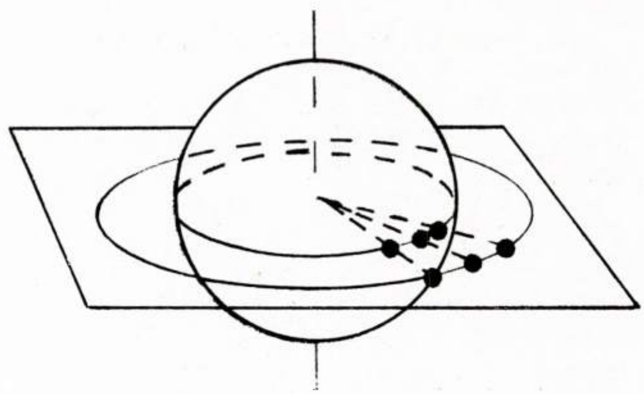


Figure 24. Equatorial track.

If the plane of the orbit is inclined to the equatorial plane, the ground track moves north and south of the equator. It moves between the limits of latitude equal to the inclination of its orbital plane (Fig. 25). A satellite in either circular or elliptical orbit will trace out a path over the earth between these same limits of latitude, determined by the inclination angle. However, the satellite in elliptical orbit will, with one exception, remain north or south of the equator for unequal periods of time. This exception occurs when the major or long axis of the orbit lies in the equatorial plane.

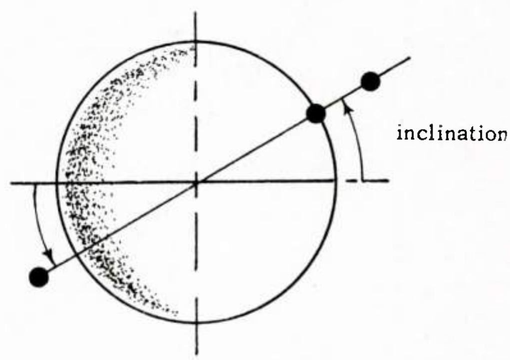


Figure 25. North-South travel limits.

The inclination of an orbit is determined by both the latitude of the vehicle and the direction of the vehicle's velocity at the time of injection or entry into orbit. That is, the cosine of the inclination angle equals the cosine of the latitude times the sine of the azimuth (when the azimuth is measured from north). The minimum inclination which an orbital plane can be made to assume is the number of degrees of latitude at which injection occurs. This minimum inclination occurs when the direction of the vehicle's velocity is due east or west at the time of injection. If the vehicle's direction at injection into orbit is any direction other than east or west, the inclination of the orbital plane will be increased (Fig. 26).

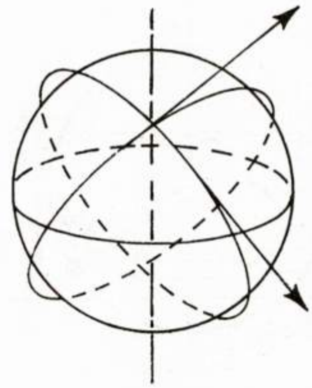


Figure 26. Injection-inclination geometry.

On a flat map of the earth, satellite ground tracks appear to have different shapes than on a sphere. The ground track for a vehicle in an inclined circular or elliptical orbit appears as a sinusoidal trace with North-South limits equal to the inclination of the orbital plane (Fig. 27).

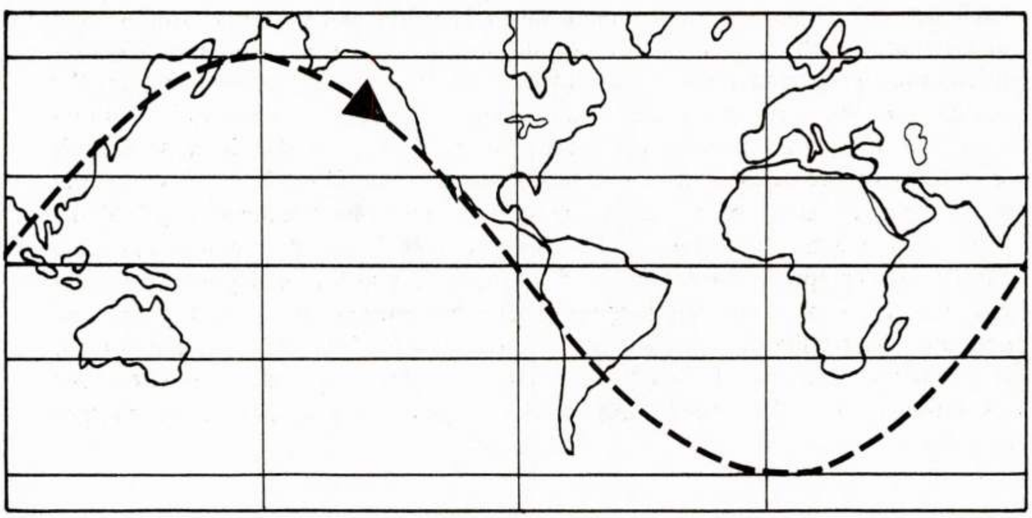


Figure 27. Ground track on flat, non-rotating earth map.

When the earth's rotation is considered, visualizing a satellite's ground track becomes more complex. A point on the equator moves from west to east more rapidly than do points north and south of the equator. Their speeds are the speed of a point on the equator times the cosine of their latitude. Satellites in circular orbit travel at a constant speed. However, when the orbits are inclined to the equator, the component of satellite velocity which is effective in an easterly or westerly direction varies continously throughout the orbital trace (Fig. 28). As the satellite crosses the equator, its easterly or westerly component of velocity is its instantaneous total velocity times the cosine of its angle of inclination.

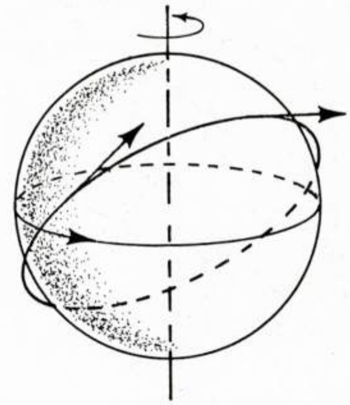


Figure 28. Effective East/West component of satellite velocity.

When it is at the most northerly or southerly portion of its orbit, its easterly or westerly component is equal to its total instantaneous velocity.

In elliptical orbits only the horizontal velocity component contributes to the satellite's ground track. Further complication results because the inertial or absolute speed of the satellite varies throughout the elliptical path (Fig. 29).

Because the ground track is dependent upon the relative motion between the satellite and the earth, the visualization of ground tracks becomes quite complicated. Earth rotation causes each successive track of a satellite in a near earth orbit to cross the equator west of the preceding track (Fig. 30). This westerly regression is equal to the period of the satellite times the rotational speed of the earth. The regression is more clearly seen if angular speed is considered. The earth's angular speed of rotation is 15°/hour. The number of degrees of regression (in terms of a shift in longitude) can be determined by multiplying the period of the satellite by 15°/hour, the angular speed of the earth. If the altitude of a satellite is increased, thereby increasing the time required to complete one revolution in the orbit, the distance between successive crossings of the equator increases.

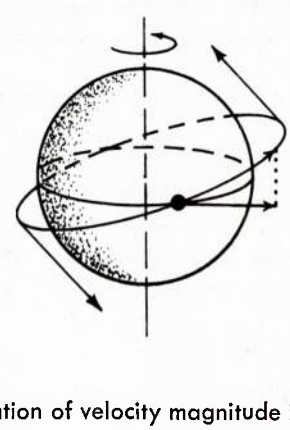


Figure 29. Variation of velocity magnitude in an elliptical orbit.

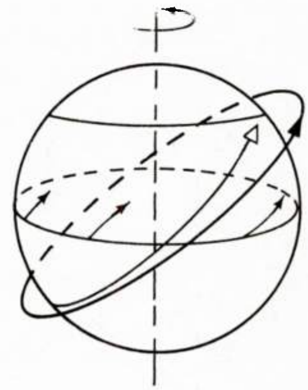


Figure 30. Ground track regression due to earth rotation.

Again, using a flat map of the earth, the track of a satellite in circular orbit (with a period of less than 20 hrs) appears as a series of sinusoidal traces, each successively displaced to the west (Fig. 31).

The ground track of a satellite in elliptical orbit results in a series of irregular traces on a flat map which have one lobe larger than the other. The lobes are compressed by an amount which depends on orbital time, are altered in shape by the combined factors (inclination, eccentricity, period, and location of perigee), and successive traces are displaced to the west (Fig. 32).

Some satellites follow orbits that have particularly interesting ground tracks. A satellite with a 24-hour period of revolution is one such case. If this satellite is in a circular orbit in the equatorial plane, it is often referred to as a *synchronous*

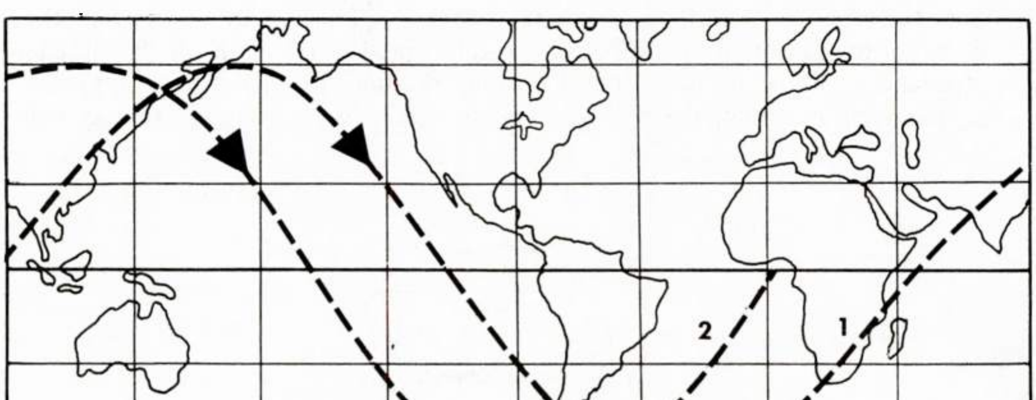


Figure 31. Westward regression of sinusoidal tracks.

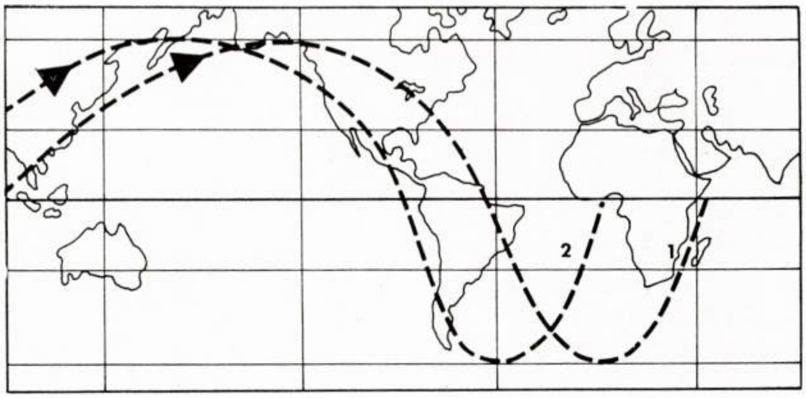


Figure 32. Westward regression of irregular tracks.

satellite; its trace is a single point. If it orbits in the polar plane, it will complete half of its orbit while the earth is rotating halfway about its axis. The result is a trace which crosses a single point on the equator as the satellite crosses the equator heading north and south. The complete ground track forms a figure eight. If the plane of the circular orbit is inclined at smaller angles to the equator, the figure eights are correspondingly smaller (Fig. 33).

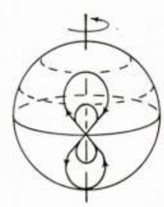


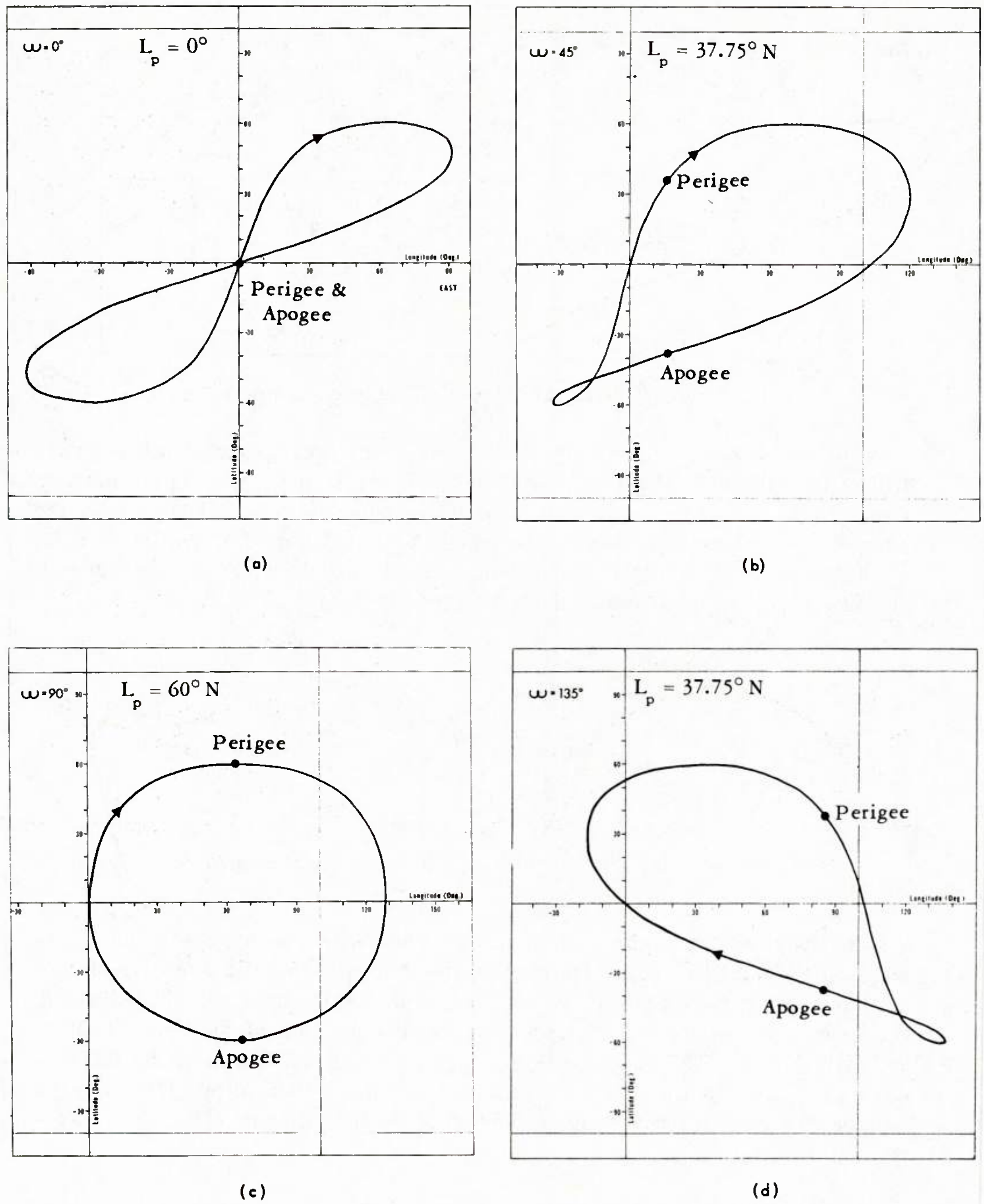
Figure 33. Figure eights for inclinations less than polar.

The shape of the figure eight may be altered by placing a satellite in an elliptical flight path. The eccentricity of the ellipse changes the relative size of the loops of the figure eight. If eccentricity and inclination are fixed, then changing the location of perigee will vary the shape and orientation of the figure eight as shown in Fig. 34. The longitude of perigee is, of course, determined by the conditions and geographical location of injection into the 24-hour orbit. However, the latitude of perigee is fixed by the inclination of the orbital plane (i), and the argument of perigee (ω):

$$\text{Sin (latitude of perigee)} = \sin i \sin \omega$$

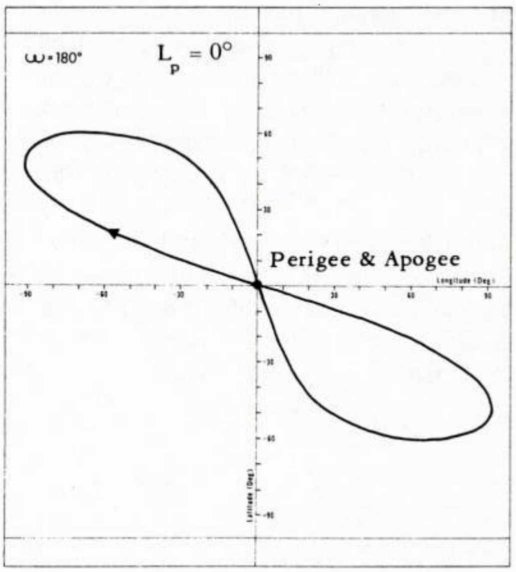
Apogee is located on the same meridian as perigee and is at the same degree of latitude but in the opposite hemisphere.

Circular ground tracks similar to the one shown in Fig. 34(g) have been proposed in certain navigation satellite concepts. In this case with apogee in the northern hemisphere, a multiple satellite system using this type of ground track might be used for supersonic aircraft navigation in the North Atlantic.

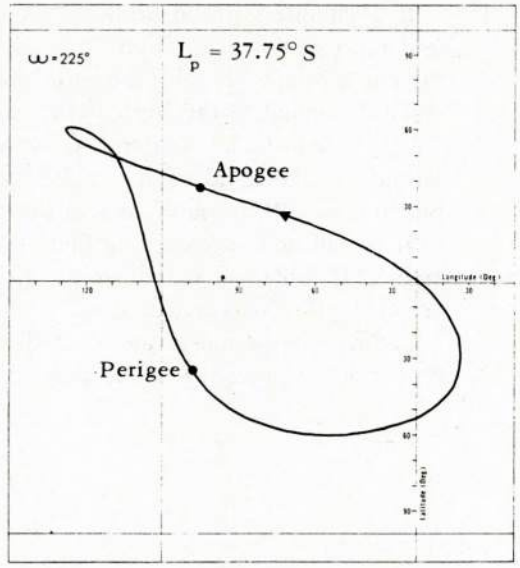


Eccentricity (ϵ) = .6
Inclination (i) = 60°
Argument of Perigee (ω) = As Stated

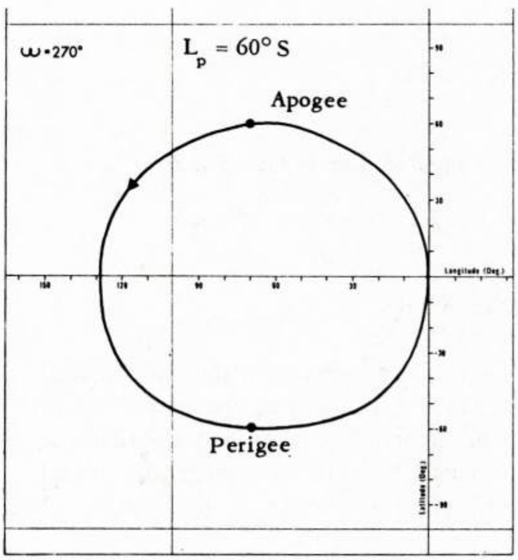
Figure 34. Variation of elliptical, 24-hour track with movement of perigee.



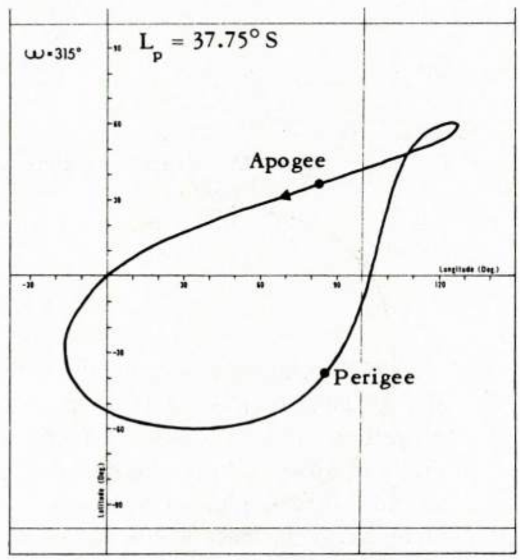
(e)



(f)



(g)



(h)

Eccentricity (ϵ) = .6
Inclination (i) = 60°
Argument of Perigee (ω) = As Stated

Figure 34. Variation of elliptical, 24-hour track with movement of perigee, continued.

If a satellite is in an orbit with a period much greater than the earth's twenty four hour period of rotation, the satellite appears as a point in space under which the earth rotates. If the orbit is in the equatorial plane, the trace is a line on the equator moving to the west. If the orbit is inclined to the equator, its earth track will appear to be a continuous trace wound around the earth like a spiral between latitude limits equal to its angle of inclination (Fig. 35). An example of this phenomena is the ground track of the moon.

It should now be clear that there is an almost limitless variety of satellite ground tracks. To obtain a particular track, it is only necessary that the proper orbit be selected. If changes are made in the inclination of an orbital plane to the equator, if its period is varied, if the eccentricity is controlled, or, if the location or perigee is specified, many different ground tracks can be achieved.

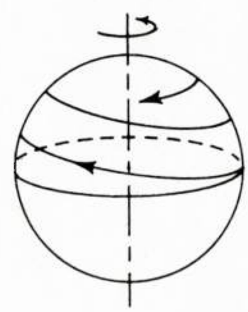


Figure 35. Track for inclined orbit with period greater than one day.

SPACE MANEUVERS

One characteristic of satellites is that their orbits are basically stable in inertial space. This stability is often an advantage, but it can also pose problems. Space operations such as resupply, rendezvous, and interception may require that the orbits of space vehicles be changed. Such changes are usually changes in orbital altitude, orbital plane (inclination), or both. In this section some methods of maneuvering in space will be reviewed.

Altitude Change

When a satellite or space vehicle is to have its orbit changed in altitude, additional energy is required. This is true whether the altitude is increased or decreased. The classic example of changing the orbital altitude of a satellite is the HOHMANN TRANSFER.

The Hohmann transfer is a two-impulse maneuver between two circular, coplanar orbits. For most practical problems, this method uses the least amount of fuel and is known as a minimum energy transfer. The path of the transfer follows an ellipse which is cotangential to the two circular orbits (Fig. 36).

In accomplishing a Hohmann transfer, two applications of thrust are required. Each application of thrust changes the speed of the vehicle and places it into a new orbit. Obviously both the direction and magnitude of the velocity change, Δv,

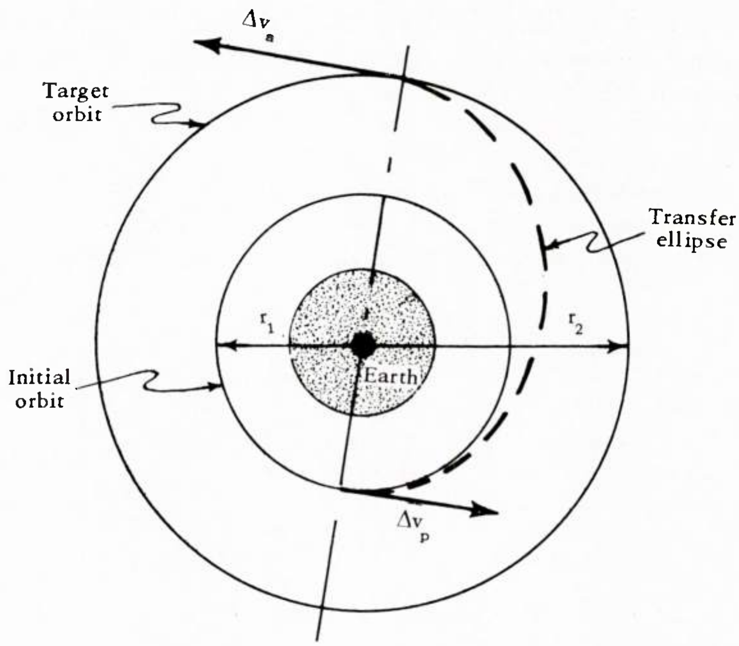


Figure 36

must be accurately controlled for a precise maneuver. If an increase in altitude is desired, the point of departure becomes the perigee of the transfer ellipse; the point of injection to the higher circular orbit becomes the apogee of the transfer ellipse. (For the transfer ellipse, $2a = r_1 + r_2$.) To lower altitude, the reverse is true. The point of departure will be the apogee of the transfer ellipse.

In general, the process for determining the total increment of velocity, Δv, required to complete a Hohmann transfer can be divided into seven steps.

(1) Determine the velocity the vehicle has in the initial orbit.

$$V_{c1} = \sqrt{\frac{\mu}{r_1}}$$

(2) Determine the velocity required at the initial point in the transfer orbit.

$$v_1 = \sqrt{\frac{2\mu}{r_1} - \frac{\mu}{a}}$$

(3) Solve for the vector difference between the velocities found in steps 1 and 2.

(4) Find the velocity the vehicle has at the final point in the transfer ellipse.

$$v_2 = \sqrt{\frac{2\mu}{r_2} - \frac{\mu}{a}} \text{ or } v_2 = \frac{v_1 r_1}{r_2}$$

(5) Compute the velocity required to keep the vehicle in the final orbit.

$$V_{c2} = \sqrt{\frac{\mu}{r_2}}$$

(6) Find the vector difference between the velocities found in steps 4 and 5.

(7) Find the total Δv for the maneuver by adding the Δv from step 3 to the Δv from step 6.

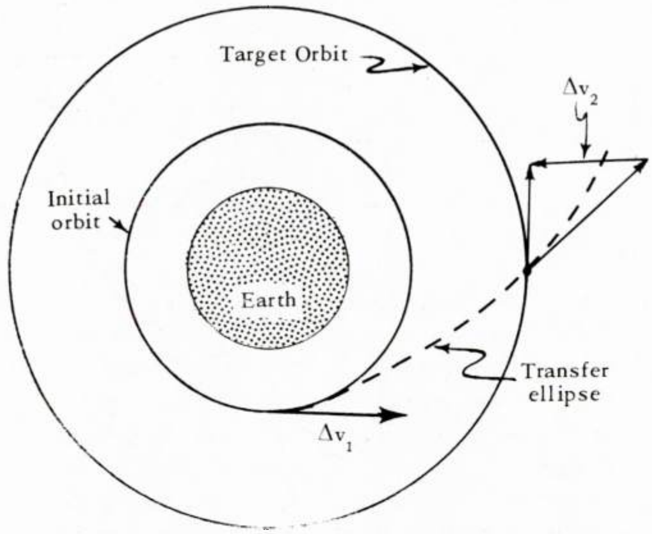


Figure 37

From a practical point, once the required Δv is known, the amount of propellant required for the maneuver can be computed from

$$\frac{\Delta v}{I_{sp}\, g} = \ln\left(\frac{W_o}{W}\right) = \ln \text{ mass ratio}$$

as discussed in Chapter 3.

There are, of course, other ways to accomplish an altitude change. One such method is the Fast Transfer, useful when time is a factor.

In the Fast Transfer, the transfer ellipse is not cotangential to the final orbit but crosses it at an angle (Fig. 37). Again the Δv is applied in two increments, but Δv_2, applied at the intersection of the transfer ellipse and the target orbit, must achieve the desired final velocity *in the proper direction.* The steps for calculating the required Δv are similar to those for the Hohmann transfer, noting that velocity differences must be treated as vectorial quantities. For the same altitude change the fast transfer requires more Δv.

Other methods of changing altitudes are discussed in other texts on astronautics. The procedures are similar to those discussed here. The magnitude and direction of the velocity vector remains the critical factor.

Plane Change

Changing the orbital planc of a satellite also requires the expenditure of energy. This is apparent if the vector diagram representing two circular orbits of the same altitude is examined, the only difference being in their inclinations:

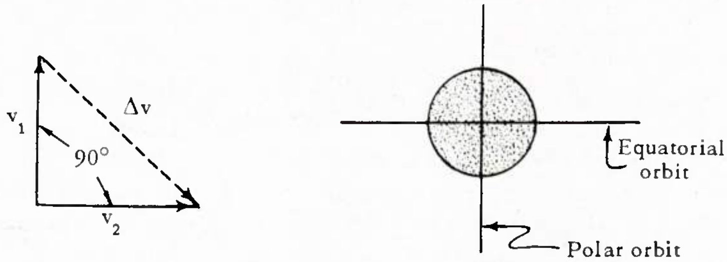
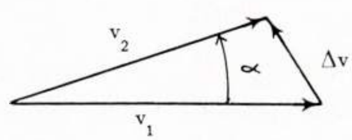


Figure 38

In this case, the orbital speeds, v_1 and v_2, are equal except that they are 90° to each other. Transferring from the polar orbit to the equatorial orbit would require the Δv represented by the dashed arrow. For a 90° plane change (an extreme case) the Δv exceeds the existing orbital speed. It should be noted that *a change from one plane to another can only be accomplished at the intersection of the two planes.*

The Δv required to change the orbital plane any specified amount can be determined by examining the vectors involved. The problem is solvable by use of the Law of Cosines. For example, if v_1 represents the existing orbital velocity, v_2 the final orbital velocity, and α the desired plane change angle,* then:



$$\Delta v^2 = v_1^2 + v_2^2 - 2v_1 v_2 \cos \alpha$$

Figure 39

If only the plane is to be changed, then $v_1 = v_2$ and the problem is simplified. But, if altitudc (or eccentricity) is also to be changed, v_1 and v_2 will not be equal.

The amount of Δv required to accomplish a plane changc is, of course, dependent upon the amount of change desired. It is also a function of the altitude at which the change is made.

* See appendix B.

Less Δv is required to make a plane change at high altitudes than at low altitudes because the orbital speed of the vehicle is less at higher altitudes. In other words, it is more economical, in terms of propellant required, to make plane changes where the speed of the satellite is low—at apogee, or at high altitudes.

Combined Maneuvers

If a requirement exists to perform both a plane change and an altitude change, some economy will result if the operations are combined.

The problem of combining a plane and altitude change is solved quite simply by considering the vector diagram. For example, it is desired to change the altitude of a vehicle from 100 NM circular orbit to 1500 NM circular orbit with a plane change of 10°. Recognizing that the plane change is more economically made at altitude, the plan is to combine the plane change with injection from the Hohmann transfer ellipse into the 1500 NM circular orbit. A typical Hohmann altitude change is initiated at a point of intersection of the two planes by increasing the vehicle's speed.

At the apogee of the transfer ellipse (1500 NM altitude) the vehicle's speed is 19,800 ft/sec. The required circular speed is 21,650 ft/sec. The complete problem, at apogee, looks like this:

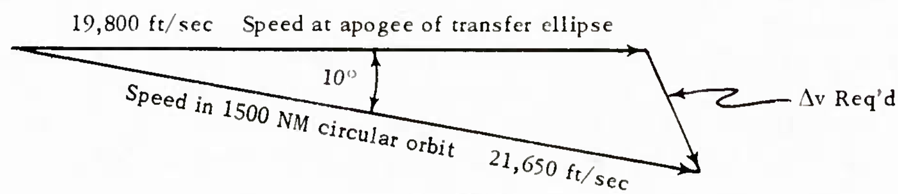


Figure 40

The Δv required for the combined maneuver is calculated by use of the Law of Cosines and is $\Delta v = 4{,}055$ ft/sec. It is also necessary to compute the angle at which the Δv is to be applied. This calculation may be accomplished with the Law of Sines.

PERTURBATIONS

When the orbit of an artificial satellite is calculated by use of the assumptions given thus far, it will vary slightly from the actual orbit unless corrections are made to take care of outside forces. These outside forces, known as perturbations, cause deviations in the orbit from those predicted by two-body orbital mechanics. In order to get a better understanding of how the satellite's actual orbit is going to behave, one must consider the following additional factors:

(1) The earth is not the only source of gravitational attraction on the satellite since there are other gravitational fields (principally of the sun and moon) in space. This effect is greatest at high altitudes (above 20,000 NM).

(2) The earth is not a spherically homogeneous mass but has a bulge around the equatorial region. This additional mass causes the gravitational pull on the satellite *not* to be directed toward the center of the earth. This is the major perturbation effect at medium altitudes (between 300 NM and 20,000 NM).

(3) The earth has an atmosphere which causes drag. This effect is most significant at low altitudes (below 300 NM).

Third Body Effects

One cause of perturbations is the introduction of one or more additional bodies to the system creating a problem involving three or more bodies. Figure 41 shows an exaggerated perturbation, referred to as a hyperbolic encounter. Initially, the satellite is in orbit 1 about an attracting body such as the earth.

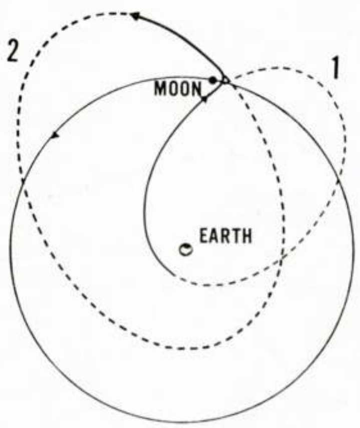


Figure 41. Hyperbolic encounter.

As the satellite approaches the moon, the gravitational influence of the moon dominates, and the center of the moon becomes the focus instead of the center of the earth. Since the vehicle approaches the moon with more than escape velocity, it must leave the moon's sphere of gravitational influence with greater than escape velocity. This means that the vehicle's velocity with respect to the moon is greater than that required to escape. Therefore, for the short time that the moon is the attracting body, the vehicle is on a hyperbolic path with respect to the moon. When the satellite leaves the sphere of influence of the moon, it switches back to the earth's sphere of influence and goes into a new elliptical orbit about the earth. The next time the satellite returns to this region, the moon will have moved in its orbit, and the satellite, therefore, will now maintain orbit 2. A hyperbolic encounter is a method of changing the energy level of a satellite. By proper positioning, it could be used to increase or decrease the energy of a space vehicle. Of course, the change in energy of the vehicle is offset by the change in energy of the second body (the moon in the case illustrated).

Effects of Oblate Earth

Another cause of perturbations is the bulge of the earth at the equator sometimes called the carth's oblateness. The effect of this oblateness on the satellite can be seen if we imagine the earth to be made up of a sphere which has an added belt of mass wrapped around the equatorial region. As shown in Fig. 42, the primary gravitational attraction F, directed to the center of the earth, will now be "disturbed" by the much smaller but nevertheless significant attractions F_1 and F_2 directed toward the near and far sides of the equatorial bulge. With r_1 smaller than r_2, F_1 will be larger than F_2, and the resultant force obtained by combining F, F_1, and F_2 will now no longer point to the center of the earth but will be deflected slightly toward the equator on the near side. As the satellite moves in its orbit the amount of this deflection will change depending on the vehicle's relative position and proximity to the equatorial region.

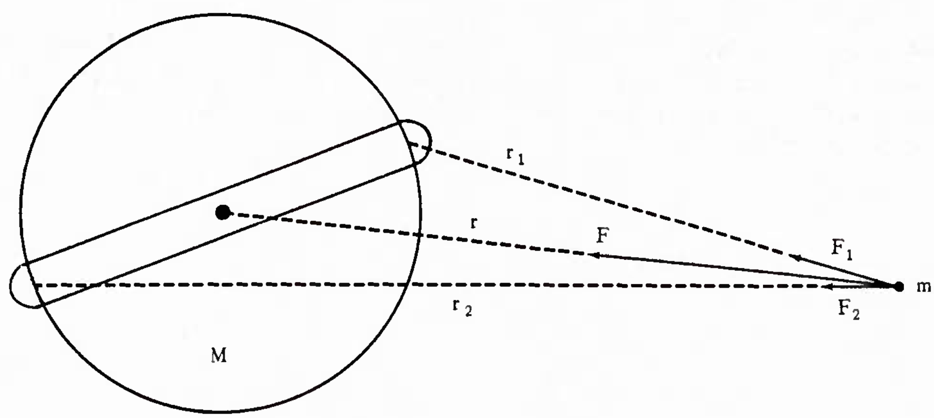


Figure 42. Deviation in force vector caused by the oblateness of the earth

Two perturbations which result from this shift in the gravitational force are:

(1) Regression of the nodes.
(2) Rotation of the line of apsides (major axis) or rotation of perigee.

Regression of the nodes is illustrated in Fig. 43 as a rotation of the plane of the orbit in spacc. The resulting effect is that the nodes, both ascending and descending, move west or east along the equator with each succeeding pass. The direction of this movement will be opposite to the east or west component of the satellite's motion. Satellites in the posigrade orbit (inclinations less than 90°) illustrated in Fig. 43 have easterly components of velocity so that the nodal regression in this case is to the west. The movement of the nodes will be reversed for retrograde orbits since they always have a westerly component of velocity. Figure 44 shows why, for a vehicle traveling west to east, the regression of the nodcs is toward the west. The original track from A to B would cross the equator at Ω_1. Simplifying the effect of the equatorial bulge to a single

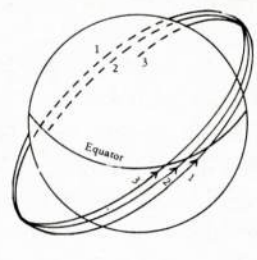


Figure 43. Regression of the nodes.

impulse at point E, the track is moved so that it crosses the equator at Ω_2. At point F the simplified effect of the bulge is a single impulse down, changing the orbital path line to the line FD, which would have crossed the equator at Ω_3. This effect, regression of the nodes, is more pronounced on low-altitude satellites than high-altitude satellites. In low altitude, low inclination orbits the regression rate may be as high as 9° per day. Fig. 45 shows how the regression rate changes for circular orbits at various altitudes and inclination angles. Note that nodal regression is zero in thc polar orbit case. It has no meaning in equatorial orbits.

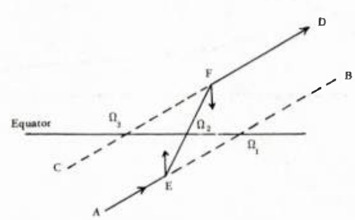


Figure 44. Regression of nodes toward the west when vehicle is traveling west to east.

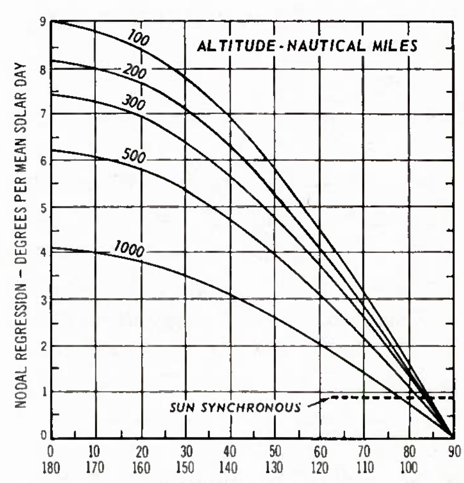


Figure 45. Nodal regression rate per day for circular orbits.

Satellites requiring sun synchronous orbits (for photography or other reasons) are an example of how regression of the nodes can be used to practical advantage in certain situations. In Fig. 46 the satellite is injected into an orbit passing over the equator on the sunlit side of the earth at local noon (the sun overhead). This condition initially aligns the orbital plane so that it contains a linc between the earth and sun. The altitude of the near circular orbit determines the angle of inclination required in order to maintain the sun synchronous nodal regression rate of approximately one degree per day (360°/year). In Fig. 45 note the required inclination angles of approximately 95° to 105° for the altitudes shown. If inclination angle and orbital altitude have been chosen correctly, then regression of the nodes will rotate the plane of the orbit (change the angle of right ascension) through 90° every three months as shown in Fig. 46. Thus, we see that this perturbing phenomenon due to the earth's equatorial bulge maintains the desired angle of right ascension which, along with the angle of inclination, orients the orbital plane within the celestial sphere. Therefore, as the earth moves in its orbit the desired orientation for best photography is maintained without using propellant.

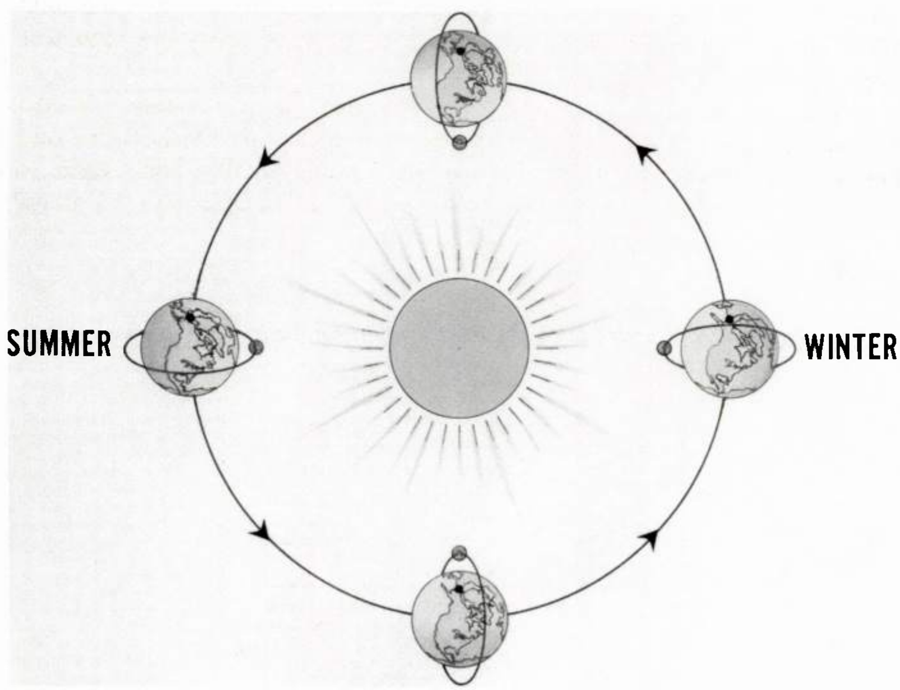


Figure 46. Sun-synchronous orbit.

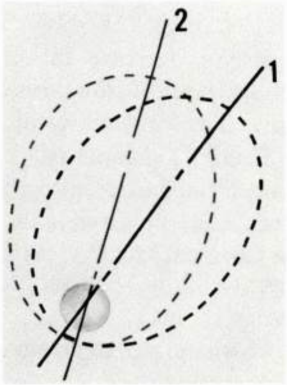


Figure 47. Earth's equatorial bulge changes the argument of perigee.

Rotation of the apsidal line is shown in Figure 47. (The apsidal line is the line joining apogee and perigee—the major axis.) The cause of this perturbation due to oblateness is difficult to visualize; however, the result is that the trajectory rotates within the orbital plane about the occupied focus. This rotation has the effect of shifting the location of perigee, thus changing the argument of perigee (Figure 21).

This rate of change in the argument of perigee is a function of satellite altitude and inclination angle. At inclinations of 63.4° and 116.6° the rate of rotation is zero. Figure 48 illustrates how the apsidal rotation rate varies with inclination angle for orbits with a 100 NM perigee and different apogee altitudes.

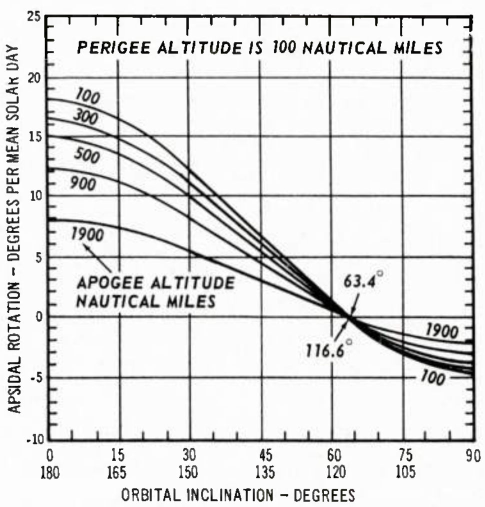


Figure 48. Apsidal rotation rate per day for orbits with 100 NM perigee altitude.

Drag Effects

Drag on a satellite will cause a decrease in eccentricity, a decrease in the major axis, and a rotation of the apsidal line. Figure 49 illustrates the change in eccentricity and major axis. The original orbit is represented by the curve A_1 A_2 A_1, with the focus at F, and a major axis equal to line segment A_1 A_2, Assuming that the drag is concentrated near perigee, the speed at A_1 eventually will diminish to circular speed, and the new path will be A_1 A_3 A_1, with the major axis decreasing to line segment A_1 A_3. After the orbit has decayed to approximately zero eccentricity (a circle), further decay will result in a nearly circular spiral with ever-decreasing radius.

Some other causes of perturbations are electromagnetic forces, radiation pressures, solar pressures, and gas-dynamic forces.

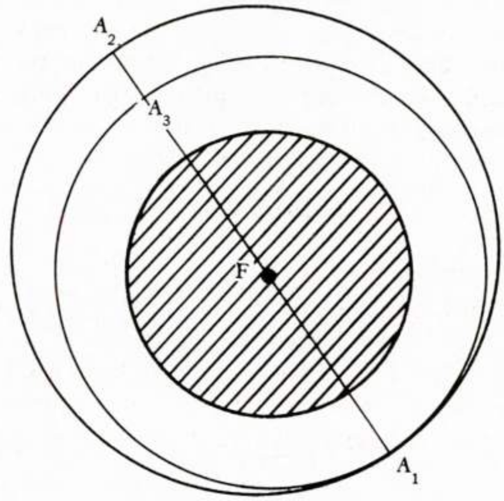


Figure 49. Decrease in the eccentricity of a satellite orbit caused by drag.

THE DEORBITING PROBLEM

The general operation of moving a body from an earth orbit to a precise point on the surface of the earth is a difficult problem. The orbit may be circular or elliptical, high or low, and inclined at various degrees to the rotational axis of the earth. The deorbiting maneuver may be the lofted, depressed, or retro approach which is the only method discussed in this section. Time of flight is a major parameter in deorbiting and presents one of the more difficult of mathematical problems. But, the theory developed here is sufficiently general in scope that it may be modified to solve most problems. The complexity of the problem is reduced in the following simplified illustration of deorbiting from a circular, polar orbit from a point over the North Pole, by assuming no atmosphere.* The approach and solution will permit impacting any earth target, provided fuel for retrothrust is no limitation.

* The student may consider the effect of the atmosphere by using the radius of reentry rather than the radius of the earth, and by then considering the range and time of flight for a specific reentry body. The solution may also be modified to consider elliptical orbits and inclined orbits.

Consider a target at 60° north latitude. If the earth could be stopped from rotating (see Fig. 50)*, with the target in the plane of the circular orbit, then it only would be necessary to apply a retrovelocity (Δv_1) to the orbital body so it would traverse the ellipse shown (arc 1-4) and impact the target. The reentry velocity magnitude (v_b), equal to the circular speed (v_c) minus the retrovelocity (Δv_1), would be necessary at apogee (1) to impact any target on the 60° north parallel. But, the direction of v_b would have to be oriented with respect to the earth's axis so that it would lie in the plane formed by the earth's axis and the predicted position of the target.

Since the earth is rotating at a constant speed, the time of flight (t_b) of the orbital object must be known in order to predict where the target will be at impact. For example, when the orbital vehicle is at Point 1 over the North Pole, the target is at Point 2. During the descent of the payload, however, the target moves to Point 3. In order to predict Point 3, t_b must be known.

Then, if the magnitudes of v_c, v_b, and α, the angle between these two vectors, are known, the Law of Cosines can be used to determine Δv_2. Application of Δv_2 to v_c insures that v_b will have the proper magnitude and direction so that the object will impact a selected target on a rotating earth.

Deorbiting Velocity

Looking first at the geometry of the problem (Fig. 51), notice that the total problem lies in one plane. Basically, the radius of the circular orbit (r_c) and the latitude of the target (L) are known. Since the original circular orbit and the bombing transfer ellipse are coplanar and cotangential, the problem begins as a Hohmann transfer. To determine Δv, the rearward velocity increment which must be applied to cause the object to impact the target can be computed from $\Delta v = v_c - v_b$, if v_c and v_b are known.

$$v_c = \sqrt{\frac{\mu}{r_c}}$$

$$v_b = \sqrt{\frac{2\mu}{r_a} - \frac{\mu}{a}}$$

$$\text{Since } r_a = a + c = a\,(1 + \epsilon),\ a_b = \frac{r_a}{1 + \epsilon_b}$$

$$\text{Then } v_b = \sqrt{\frac{2\mu}{r_a} - \mu\,\frac{(1 + \epsilon_b)}{r_a}} = \sqrt{\frac{\mu}{r_a}\,(1 - \epsilon_b)}$$

Write the general equation of the conic as $k\epsilon = r\,(1 + \epsilon \cos \nu)$ and evaluate at the two known points, r_e** and r_a, noting that at impact $\cos \nu = -\cos \theta_t = -\sin L$, and at retrofire $\cos(180°) = -1$.

Then $k\epsilon = r_e\,(1 - \epsilon_b \sin L) = r_a\,(1 - \epsilon_b)$

* 8 figures related to the deorbit problem appear at the end of this section.

** r_e is radius of the earth, but radius of re-entry could be used if desired.

Solving for ϵ_b.

$$\epsilon_b\ (r_a - r_e \sin L) = r_a - r_e$$

$$\epsilon_b = \frac{r_a - r_e}{r_a - r_e \sin L}$$

Note that in order to find ϵ_b and, in turn, the magnitude of v_b, only the *latitude* of the target and the *altitude* of the circular orbit need be known. Suppose there is a satellite in a 500 NM circular polar oribit. When the orbital vehicle arrives directly over the North Pole, it is desired to deorbit an object which will impact at 60° north latitude. Assuming that the earth has no atmosphere and is nonrotating, what retrovelocity is necessary?

$$\epsilon_b = \frac{r_a - r_e}{r_a - r_e \sin L} = \frac{(23.94 \times 10^6\,\text{ft}) - (20.9 \times 10^6\,\text{ft})}{(23.94 \times 10^6\,\text{ft}) - (20.9 \times 10^6\,\text{ft}) \quad \sin 60°}$$

$$\epsilon_b = .520$$

$$v_b = \sqrt{\frac{\mu}{r_a}(1 - \epsilon_b)} = \sqrt{\frac{14.08 \times 10^{15}\ \text{ft}^3/\text{sec}^2}{23.94 \times 10^6\ \text{ft}}(1 - .520)}$$

$$v_b \doteq 16{,}790\ \text{ft/sec}$$

$$\Delta v = v_c - v_b = 24{,}240\ \text{ft/sec} - 16{,}790\ \text{ft/sec}$$

$$\Delta v = 7{,}450\ \text{ft/sec}$$

Looking again at the geometry, note that there is a satellite in a 500 NM circular, polar orbit with $v_c = 24{,}240$ ft/sec. A retrovelocity increment, Δv, equal to 7,450 ft/sec was applied. This provided a magnitude $v_b = 16{,}790$ ft/sec so that the object now follows arc 1-2 and impacts on 60° north latitude.

Figure 52 provides a graph of velocity versus latitude to determine the magnitude of v_b. The graph is for *a specific altitude,* for release from over the North Pole, and for no atmosphere. Included also are more general graphs, Figures 53 and 54.

Deorbit Time of Flight

As the bomb falls from apogee, the target moves toward the east due to earth rotation. Its speed is 1520 cos L ft/sec, so the necessity of accurately computing the time to bomb, t_b, is readily apparent.

First, look at the problem in schematic, Fig. 50. Recall that at the time the vehicle is over the North Pole at Point 1, the target is at Point 2. Thus, t_b must be known so that the location of Point 3 can be predicted. If the meridian that Point 3 will be on, and the meridian with which the circular orbit coincides at the instant of deorbit are known, then the angle α can be determined by subtraction.

There are several methods for computing t_b from release to impact. Eqn 10, App D could be used to determine t_ψ, the time of flight for a ballistic missile launched

from 60° N latitude, with apogee at 500 NM over the North Pole, and impacting at 60° N latitude. This would be the t_ψ to traverse arc 5-1-4 (Fig 50). The actual time to bomb from apogee would be half this amount. The general time of flight method is presented here.

$$t_{1 \to 2} = \sqrt{\frac{a^3}{\mu}} (u_2 - \epsilon \sin u_2) - (u_1 - \epsilon \sin u_1) \qquad \text{(App D, Eqn 7)}$$

Since position 1 corresponds to apogee:

$$u_1 = \pi, \ \sin u_1 = 0$$

$$\text{Then } t_b = t_{1 \to 2} = \sqrt{\frac{a^3}{\mu}} (u_t - \epsilon_b \sin u_t - \pi)$$

$$a = \frac{r_a}{1 + \epsilon_b}$$

$$\cos u_t = \frac{\epsilon_b - \cos \theta_t}{1 - \epsilon \cos \theta_t} = \frac{\epsilon_b - \sin L}{1 - \epsilon_b \sin L}$$

All of these parameters are familiar except u which is the eccentric anomaly (Fig. 55). If a perpendicular is dropped through the target to the major axis of the ellipse, it intersects a circle (with the center at C, and the diameter equal to the major axis of the ellipse) at Point Q. By definition, angle BCQ is u, the eccentric anomaly. The radius of the circular orbit, r_a, is given. Both v_b and ϵ_b can be calculated from formulas given previously.

Working with the same example that was used to illustrate deorbiting velocity, t_b will be calculated from only two known quantities—*the altitude of the satellite and the latitude of the target.* A satellite is in a 500 NM circular, polar orbit. Find the time of flight, t_b, from directly over the North Pole to a target on the 60° north parallel:

$$\epsilon_b = \frac{r_a - r_e}{r_a - r_e \sin L} = \frac{(23.94 \times 10^6 \text{ ft}) - (20.9 \times 10^6 \text{ ft})}{(23.94 \times 10^6 \text{ ft}) - (20.9 \times 10^6 \text{ ft}) \sin 60°}$$

$$\epsilon_b = .520$$

$$v_b = \sqrt{\frac{\mu}{r_a} (1 - \epsilon_b)} = \sqrt{\frac{14.08 \times 10^{15} \text{ ft}^3/\text{sec}^2 (1 - .520)}{23.94 \times 10^6 \text{ ft}}}$$

$$v_b = 16{,}790 \text{ ft/sec}$$

$$a = \frac{r_a}{1 + \epsilon_b} = \frac{23.94 \times 10^6 \text{ ft}}{1.520} = 15.78 \times 10^6 \text{ ft}$$

$$\cos u_t = \frac{\epsilon_b - \sin L\,.}{1 - \epsilon_b \sin L} = \frac{.520 - \sin 60°}{1 - .520 \sin 60°} = -.637$$

Noting u_t lies between 180° and 270°:

$$u_t = 360° - 129.6° = 230.4° = 4.084 \text{ radians}$$

$$t_b = \sqrt{\frac{a^3}{\mu}} \left[u_t - \epsilon_b \sin u_t - \pi\right]$$

$$t_b = \sqrt{\frac{(15.78 \times 10^6 \text{ ft})^3}{14.08 \times 10^{15} \text{ ft}^3/\text{sec}^2}} \left[4.084 - .520 \sin 230.4° - 3.1416\right]$$

$$t_b = 676 \text{ sec} = 11.27 \text{ min}$$

Figure 56 shows t_b in minutes versus a plot of target latitude. Once again it must be recognized that this chart is for a specific orbital altitude and no atmosphere. See also the more general graph, Figure 57.

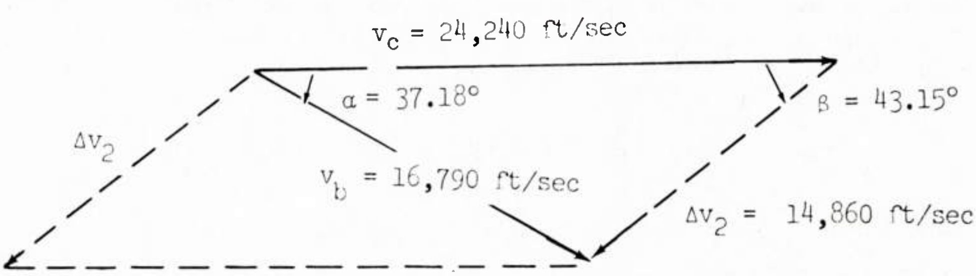
Consider impacting a target at 60° north latitude and 30° east longitude. The plane of the satellite is coincident with the 70° east meridian. This means that, when the satellite is at Point 1 and the target is at Point 2, the angle between the target and orbital plane is 40°. However, *the target is moving.* It moves at:

$$(360°/24 \text{ hrs}) \frac{\text{hr}}{60 \text{ min}} = .25 \text{ deg/min}$$

$$\theta_r = (11.27 \text{ min})(.25 \text{ deg/min}) = 2.82°$$

$$\alpha = 40° - 2.82° = 37.18°$$

Now apply the Law of Cosines and determine the Δv_2 which must be applied to v_c in order to impact the target:



$$\Delta v_2 = (v_b^2 + v_c^2 - 2 v_b v_c \cos \alpha)^{1/2}$$

$$\Delta v_2 = [(16{,}790 \text{ ft/sec})^2 + (24{,}240 \text{ ft/sec})^2 - 2(16{,}790 \text{ ft/sec} \times 24{,}240 \text{ ft/sec} \cos 37.18°)]^{1/2}$$

$$\Delta v_2 = [2.82 \times 10^8 + 5.88 \times 10^8 - 6.49 \times 10^8]^{1/2}$$

$$\Delta v_2 = [2.21 \times 10^8]^{1/2} = 14{,}860 \text{ ft/sec}$$

Also, the value of angle β must be known so that the proper direction of Δv_2 may be determined. The Law of Sines is preferred for this calculation, although the Law of Cosines can be used.

$$\sin \beta = \frac{v_b \sin \alpha}{\Delta v_2}$$

$$\sin \beta = \frac{16{,}790 \text{ ft/sec} \sin 37.18°}{14{,}860 \text{ ft/sec}}$$

$$\sin \beta = .6827$$

$$\beta = 43.15°$$

From the solution of this simplified object-from-orbit problem, it is apparent that such a calculation is not really simple. Other mathematical approaches and other operational problems are even more difficult. However, the theory presented can be extended to more difficult cases.

Fuel Requirement

It is also of interest to determine the propellant necessary to perform this maneuver. Assuming an $I_{sp} = 450$ sec (a reasonable figure in the near future) and an initial weight of 10,000 pounds, compute the amount of propellant required, W_p:

$$\ln \left(\frac{W_1}{W_2}\right) = \frac{\Delta v}{I_{sp} g} = \frac{14{,}860 \text{ ft/sec}}{(450 \text{ sec})\ (32.2 \text{ ft/sec}^2)} = 1.025$$

$$\frac{W_1}{W_2} = 2.79$$

$$W_2 = \frac{10{,}000 \text{ lbs}}{2.79} = 3{,}590 \text{ lbs}$$

$$W_p = W_1 - W_2 = 6{,}410 \text{ lbs}$$

This weight of propellant represents $\frac{W_p}{W_1} = \frac{6{,}400}{10{,}000} = 64.1\%$ of the weight in orbit prior to maneuvering.

By restricting the plane change (side range) to $\alpha = 7.18$ degrees, $\Delta v_2 = 7{,}875$ ft/sec, $W_p = 4{,}190$ lbs, and $\frac{W_p}{W_1} = 41.9\%$.

In summary, recall that the altitude of a satellite in a circular orbit and a time when the satellite was over the North Pole were given. A target was selected, and no earth atmosphere was assumed. From the theory, the required velocity for a deorbiting object was calculated, and the position of the target at the time of impact was predicted. With this information and the use of the laws of sines and cosines, the magnitude of retrovelocity and the direction to deorbit on target were calculated. Also, the amount of propellant required was computed.

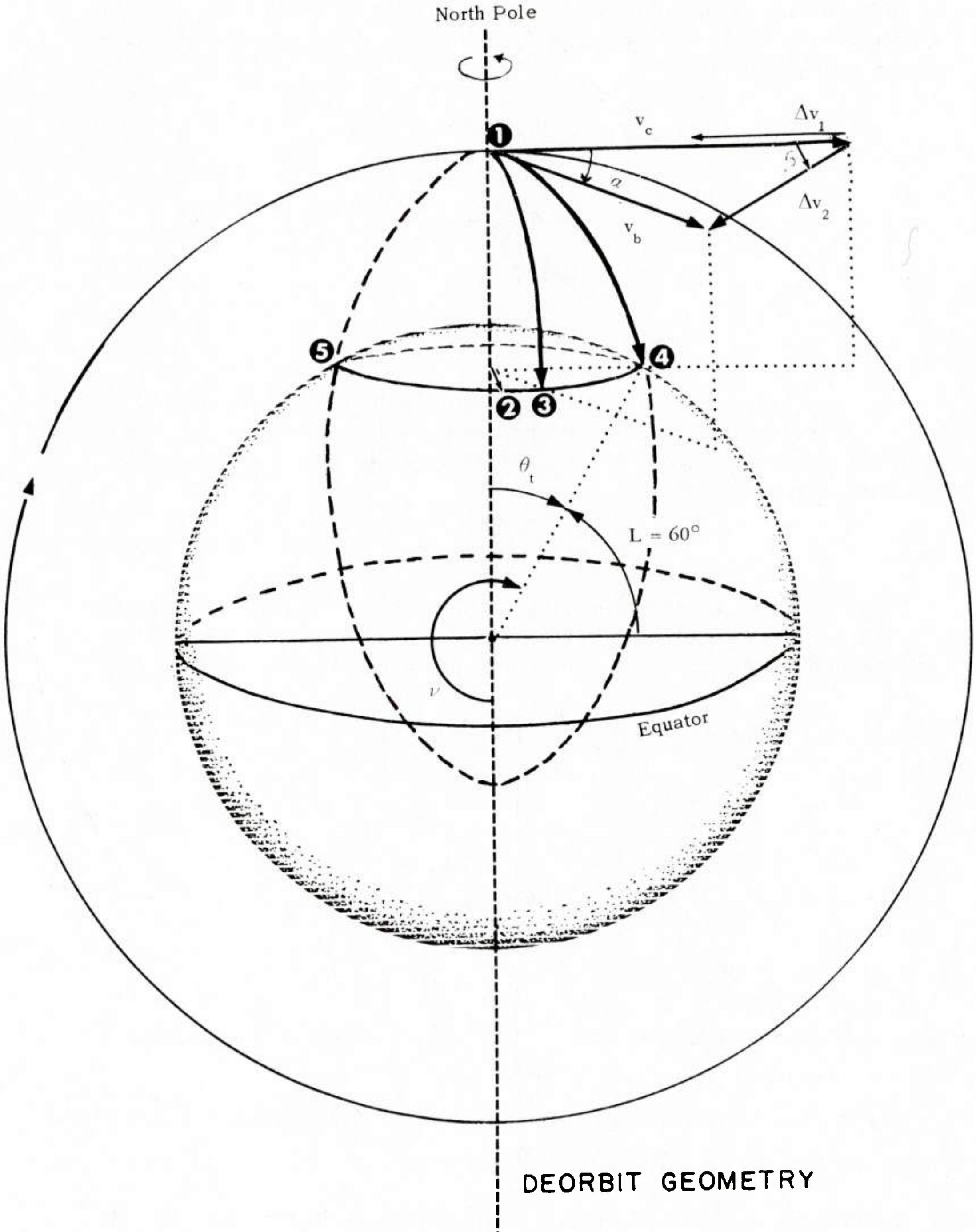

North Pole
v_c
Δv_1
β
α
Δv_2
v_b
1
2
3
4
5
θ_t
L = 60°
ν
Equator
DEORBIT GEOMETRY

Figure 50

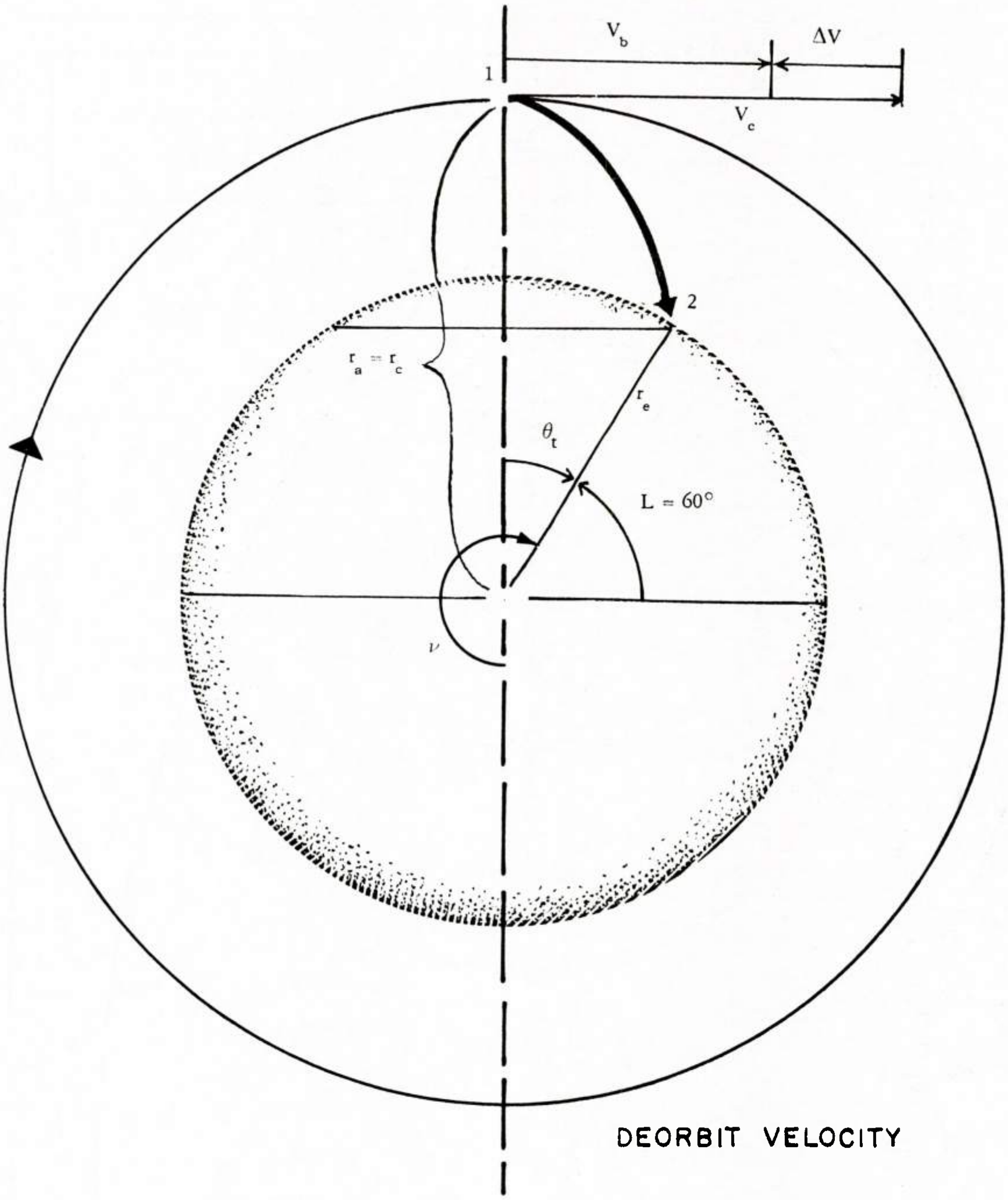


Figure 51

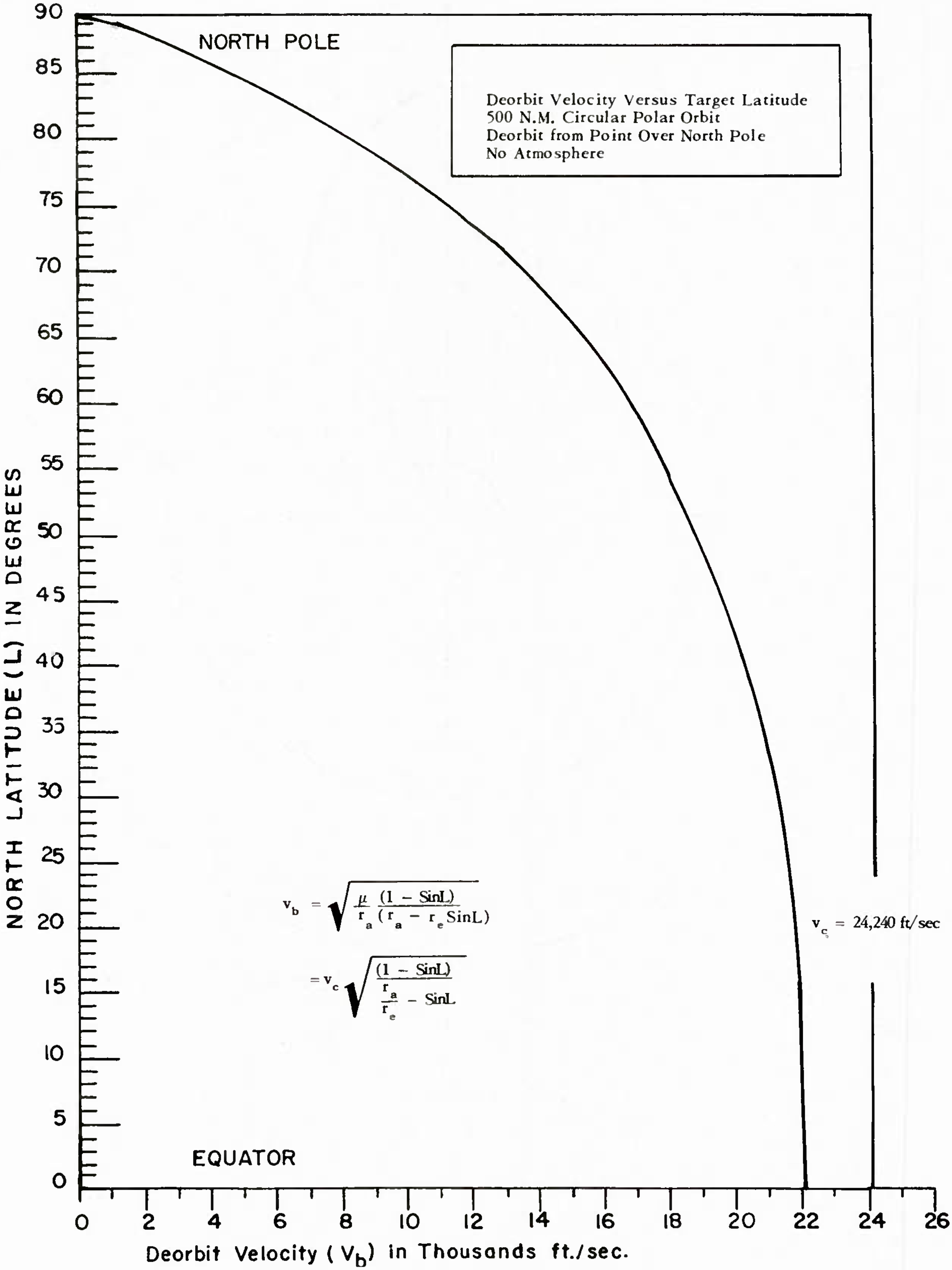

NORTH POLE
Deorbit Velocity Versus Target Latitude
500 N.M. Circular Polar Orbit
Deorbit from Point Over North Pole
No Atmosphere
$v_b = \sqrt{\frac{\mu}{r_a} \frac{(1 - \mathrm{Sin}L)}{(r_a - r_e \mathrm{Sin}L)}}$
$= v_c \sqrt{\frac{(1 - \mathrm{Sin}L)}{\frac{r_a}{r_e} - \mathrm{Sin}L}}$
v_c = 24,240 ft/sec
EQUATOR
NORTH LATITUDE (L) IN DEGREES
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
0 2 4 6 8 10 12 14 16 18 20 22 24 26
Deorbit Velocity (V_b) in Thousands ft./sec.

Figure 52

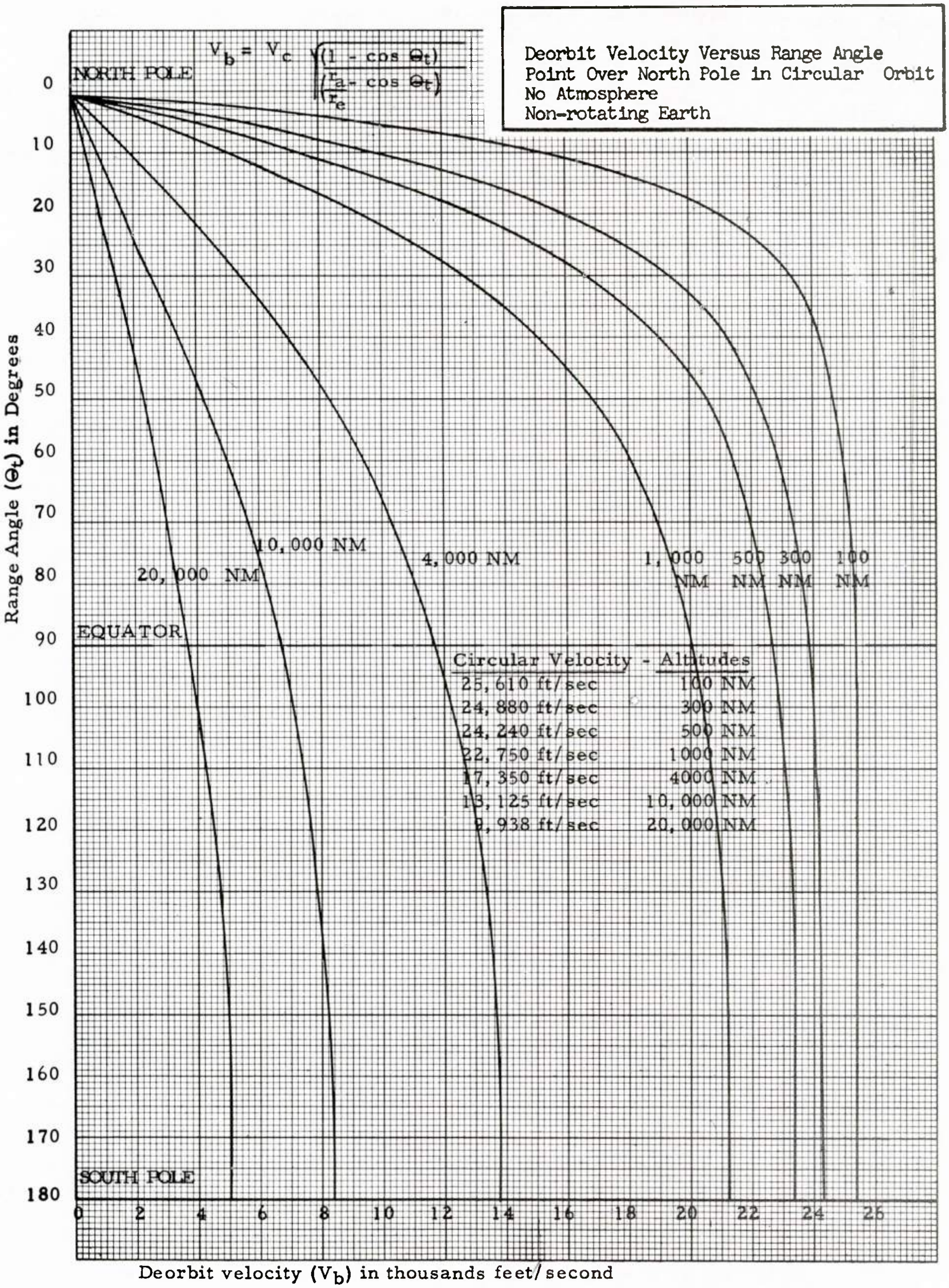

Deorbit Velocity Versus Range Angle
Point Over North Pole in Circular Orbit
No Atmosphere
Non-rotating Earth
$V_b = V_c \sqrt{\dfrac{(1 - \cos \Theta_t)}{\left(\dfrac{r_a}{r_e} - \cos \Theta_t\right)}}$
NORTH POLE
EQUATOR
SOUTH POLE
20,000 NM
10,000 NM
4,000 NM
1,000 NM
500 NM
300 NM
100 NM
Circular Velocity - Altitudes
25,610 ft/sec 100 NM
24,880 ft/sec 300 NM
24,240 ft/sec 500 NM
22,750 ft/sec 1000 NM
17,350 ft/sec 4000 NM
13,125 ft/sec 10,000 NM
9,938 ft/sec 20,000 NM
Range Angle (Θ_t) in Degrees
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
0 2 4 6 8 10 12 14 16 18 20 22 24 26
Deorbit velocity (V_b) in thousands feet/second

Figure 53

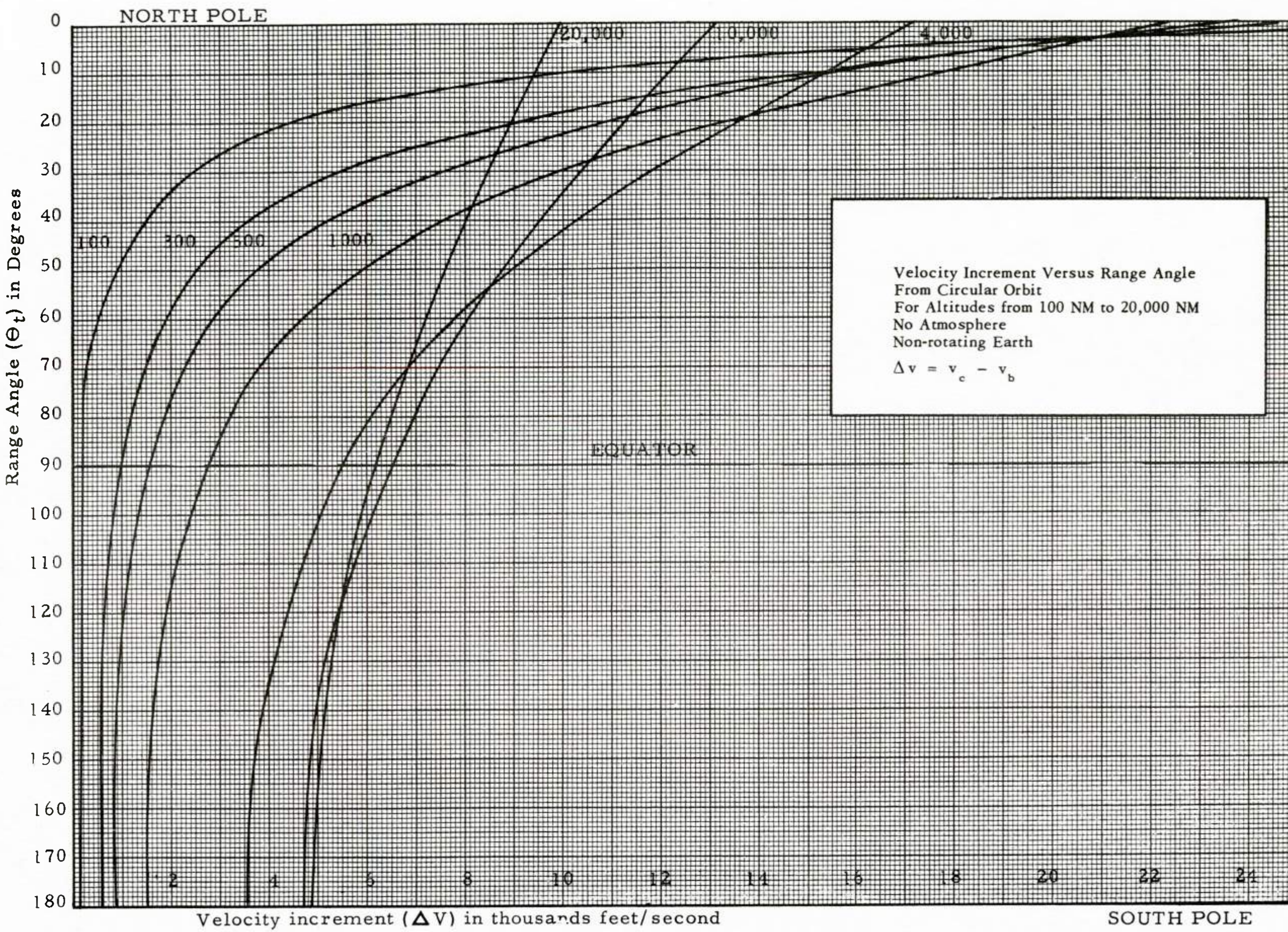

NORTH POLE
SOUTH POLE
EQUATOR
Range Angle (Θ_t) in Degrees
Velocity increment (ΔV) in thousands feet/second
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
2
4
6
8
10
12
14
16
18
20
22
24
100
300
500
1000
20,000
10,000
4,000
Velocity Increment Versus Range Angle
From Circular Orbit
For Altitudes from 100 NM to 20,000 NM
No Atmosphere
Non-rotating Earth
$\Delta v = v_c - v_b$

Figure 54

TIME OF FLIGHT GEOMETRY

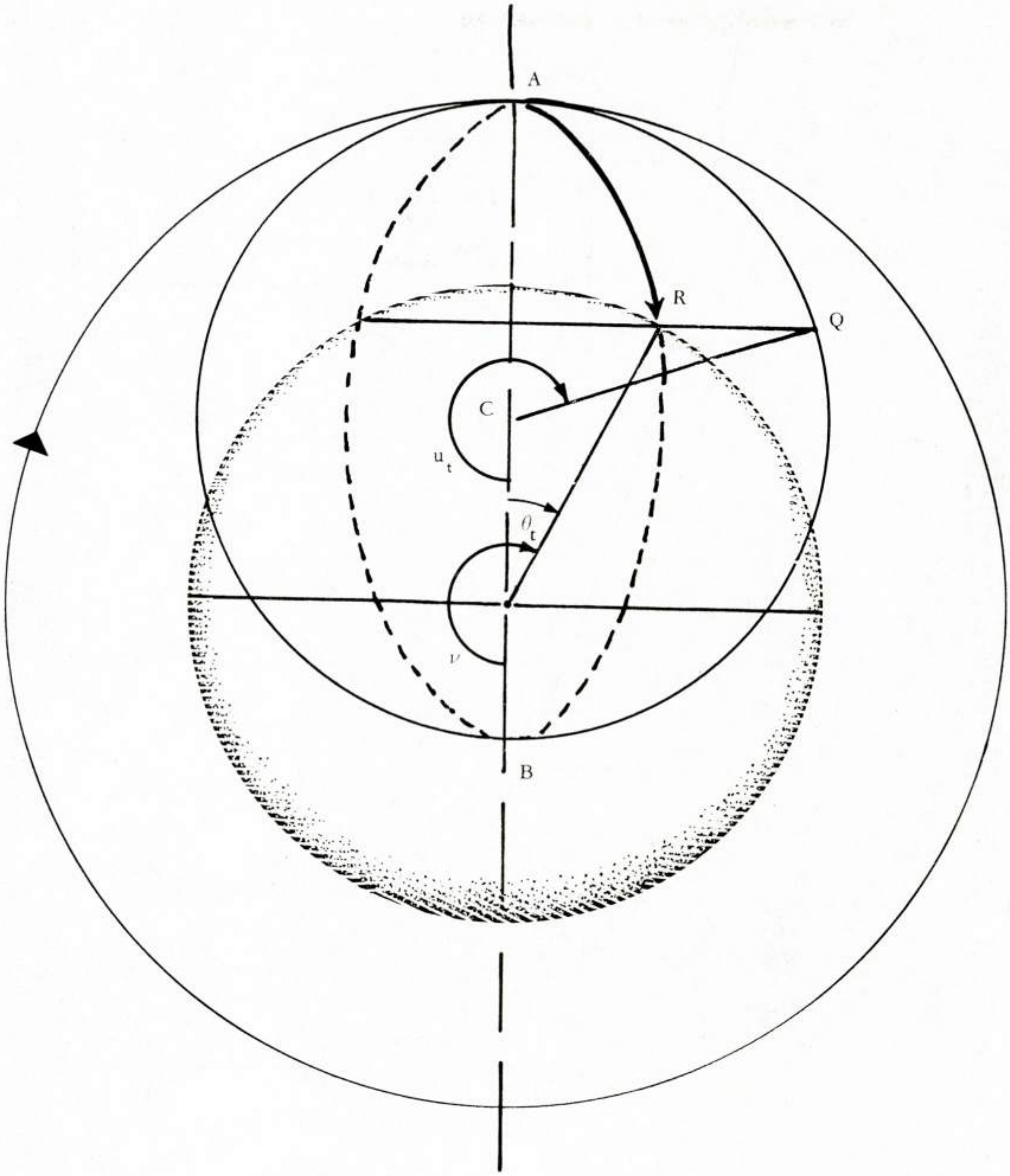
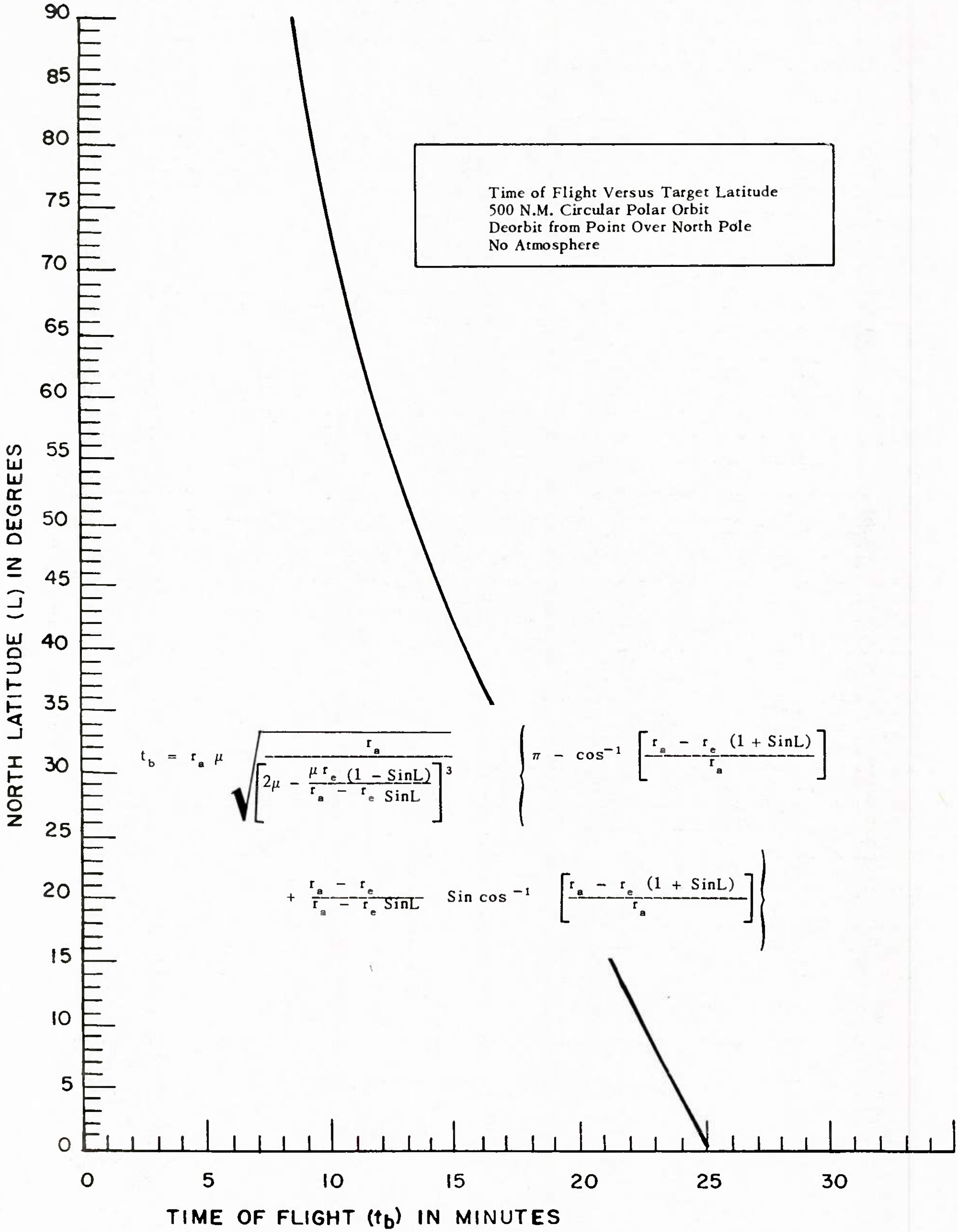


Figure 55



$$t_b = r_a \mu \sqrt{\frac{r_a}{\left[2\mu - \frac{\mu r_e (1 - \mathrm{Sin}L)}{r_a - r_e \mathrm{Sin}L}\right]^3}} \left\{ \pi - \cos^{-1}\left[\frac{r_a - r_e (1 + \mathrm{Sin}L)}{r_a}\right] + \frac{r_a - r_e}{r_a - r_e \mathrm{Sin}L} \mathrm{Sin} \cos^{-1}\left[\frac{r_a - r_e (1 + \mathrm{Sin}L)}{r_a}\right] \right\}$$

Figure 56

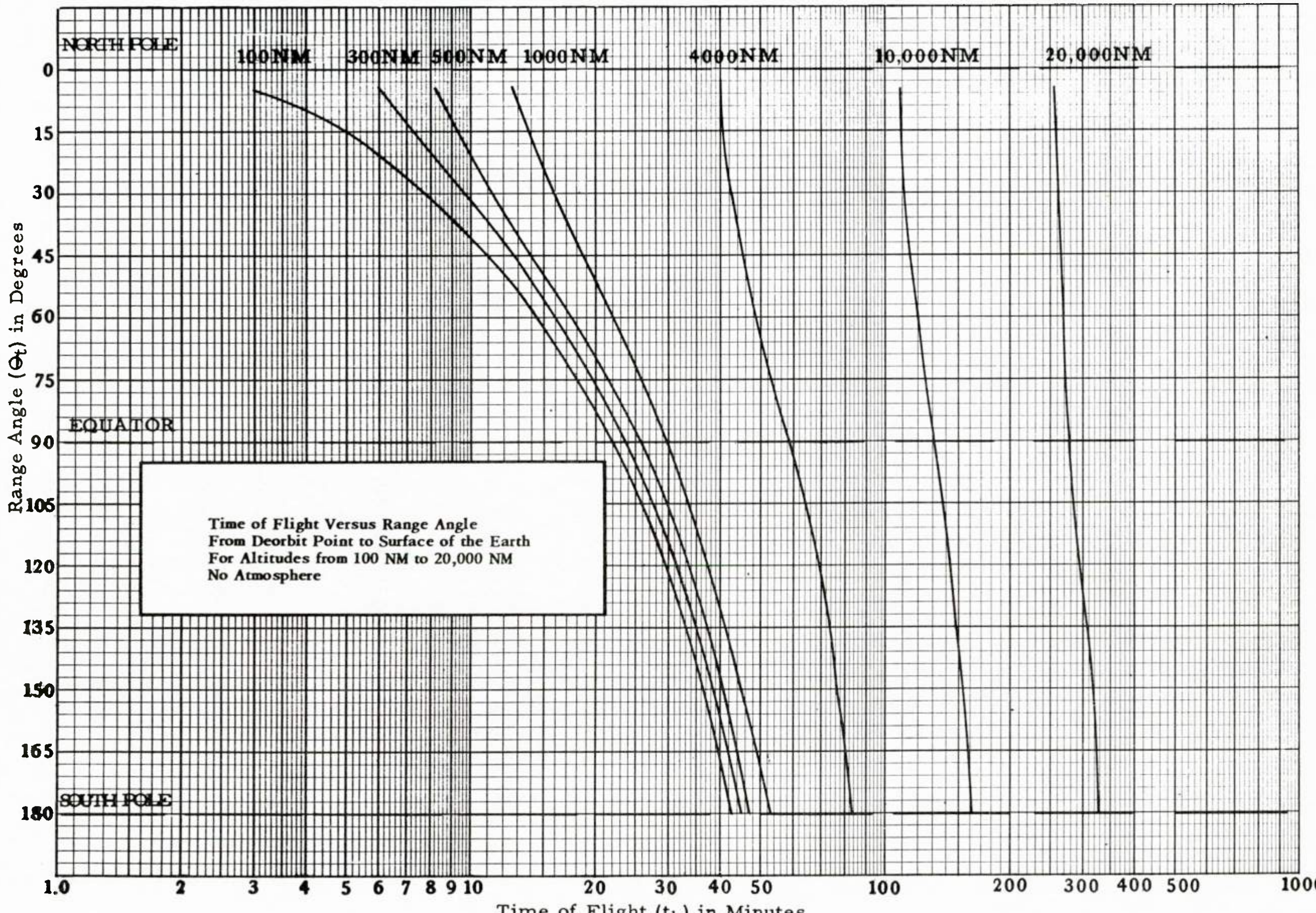

NORTH POLE
EQUATOR
SOUTH POLE
100NM
300NM
500NM
1000NM
4000NM
10,000NM
20,000NM
Time of Flight Versus Range Angle
From Deorbit Point to Surface of the Earth
For Altitudes from 100 NM to 20,000 NM
No Atmosphere
Range Angle (Θ_t) in Degrees
0
15
30
45
60
75
90
105
120
135
150
165
180
1.0
2
3
4
5
6
7
8
9
10
20
30
40
50
100
200
300
400
500
1000
Time of Flight (t_b) in Minutes

Figure 57

EQUATIONS PERTAINING TO BODIES IN MOTION

Linear Motion

1\. (a) $v_{av} = \dfrac{s_f - s_o}{t_f - t_o} = \dfrac{\Delta s}{\Delta t}$

2\. (a) $a_{av} = \dfrac{v_f - v_o}{t_f - t_o} = \dfrac{\Delta v}{\Delta t}$

for uniform linear acceleration:

3\. (a) $s = v_o t + \dfrac{at^2}{2}$

4\. (a) $v_f = v_o + at$

5\. (a) $2as = v_f^2 - v_o^2$

Angular Motion

1\. (b) $\omega_{av} = \dfrac{\theta_f - \theta_o}{t_f - t_o} = \dfrac{\Delta \theta}{\Delta t}$

2\. (b) $\alpha_{av} = \dfrac{\omega_f - \omega_o}{t_f - t_o} = \dfrac{\Delta \omega}{\Delta t}$

for uniform angular acceleration:

3\. (b) $\theta = \omega_o t + \dfrac{\alpha t^2}{2}$

4\. (b) $\omega_f = \omega_o + \alpha t$

5\. (b) $2\alpha\theta = \omega_f^2 - \omega_o^2$

Conversions from Angular Motion to Linear Motion

6\. (a) $s = r\theta$

(b) $v_t = r\omega$

(c) $a_t = r\alpha$

7\. $v_t = \dfrac{2\pi r}{P}$

8\. $a_r = \dfrac{v_t^2}{r}$

Symbols

a—linear acceleration
P—period
r—radius
s—linear displacement
t—time interval
v—linear velocity
α—angular acceleration
θ—angular displacement
ω—angular velocity

Subscripts

av—average
f—final value
o—original value
r—radial
t—tangential

SOME USEFUL EQUATIONS OF ORBITAL MECHANICS

1\. Eccentricity: $\epsilon = \dfrac{c}{a} = \dfrac{r_a - r_p}{r_a + r_p}$

2\. Ellipse:

$$r_a + r_p = 2a$$
$$a - c = r_p$$
$$a + c = r_a$$
$$a^2 = b^2 + c^2$$

3. Specific Mechanical Energy:

$$E = \frac{v^2}{2} - \frac{\mu}{r}$$

$$E = -\frac{\mu}{2a}$$

4. Specific Angular Momentum:

$$H = v\,r\cos\phi$$

5. Two Body Relationships:

$$V_{circle} = \sqrt{\frac{\mu}{r}}$$

$$V_{ellipse} = \sqrt{\frac{2\mu}{r} - \frac{\mu}{a}}$$

$$V_{escape} = \sqrt{\frac{2\mu}{r}} = \sqrt{2gr}$$

$$P = \frac{2\pi a^{\frac{3}{2}}}{\mu^{\frac{1}{2}}} = \left(5.30 \times 10^{-8} \frac{\sec}{ft^{\frac{3}{2}}}\right)(a)^{\frac{3}{2}}$$

$$P^2 = \frac{4\pi^2 a^3}{\mu} = \left(2.805 \times 10^{-15} \frac{\sec^2}{ft^3}\right)(a)^3$$

6. Variation of g:

$$g = \frac{\mu}{r^2}$$

7. Law of Cosines:

$$a^2 = b^2 + c^2 - 2bc\cos A$$

Law of Sines:

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$$

8. Constants:

$$r_e = 20.9 \times 10^6 \text{ ft}$$

$$r_e = 3440 \text{ NM}$$

$$\mu = 14.08 \times 10^{15} \frac{ft^3}{\sec^2} \quad \text{For earth.}$$

$$1 \text{ NM} = 6080 \text{ ft}$$

$$\pi \text{ radians} = 180°$$

$$1 \text{ radian} = 57.3°$$

9. Inclination of Orbital Plane

$$\cos i = \cos \text{Lat} \sin \text{Azimuth}$$

BIBLIOGRAPHY

Air Force. U.S. Air Force Academy. "Astronautics, Elements of Free Flight Trajectories." N. P.: Department of Astronautics, United States Air Force Academy, 1961. (U.S. Air Force Academy classroom text.)

Berman, Arthur I. *The Physical Principles of Astronautics.* New York: John Wiley and Sons, Inc., 1961.

Ehricke, Kraft. *Environment and Space Mechanics.* Princeton, N. J.: Van Nostrand, 1960. (Vol. I of *Space Flight*)

________. *Dynamics.* Princeton, N.J.: Van Nostrand, 1960. (Vol. II of *Space Flight*)

Ely, Lawrence D. *Space Science for the Layman.* Springfield, Illinois: Charles C. Thomas, 1967.

El'yasberg, P. E. *Introduction to the Theory of Flight of Artificial Earth Satellites.* (Translated from Russian.) Jerusalem: Israel Program for Scientific Translations, 1967.

Fowles, Grant R. *Analytical Mechanics.* Chicago: Holt, Rinehart, and Winston, 1963.

Glasstone, Samuel. *Sourcebook on the Space Sciences.* Princeton, New Jersey: D. Van Nostrand Co., Inc., 1965.

Koestler, Arthur. *The Sleepwalkers.* New York: The MacMillan Co., 1959.

LeGalley, Donald P. *Guidance, Navigation, Tracking, and Space Physics,* New York: Academic Press, 1960. (Vol. III of *Ballistic Missile and Space Technology*)

Navy, U.S. Naval Academy. "Space Flight Handbook." Washington: United States Government Printing Office, 1962. (NAVPERS 92988)

Ruppe, Harry O. *Introduction to Astronautics.* New York: Academic Press Inc., 1966.

Szebehely, Victor G. *Celestial Mechanics and Astrodynamics.* New York: Academic Press, 1964. (Vol. 14 of *Progress in Astronautics and Aeronautics*)

________. *Theory of Orbits.* New York: Academic Press, 1967.

CHAPTER 3

PROPULSION SYSTEMS

THROUGHOUT HISTORY, the methods of propulsion that man has been able to control have allowed him to go faster, farther, and higher. Today he uses rockets to send payloads into space. A space mission needs a launch vehicle or booster which will withstand high acceleration and aerodynamic forces as it lifts the payload from the surface of the earth. To be successful, the mission also requires adequate ground equipment to launch and track the flight vehicle, propulsion and guidance to place it on the desired trajectory, reliable electric power sources, and communication equipment to send data back to earth. Finally, mission success requires trained personnel who follow correct procedures to insure that all subsystems operate properly.

Rocket propulsion is vital in any successful space program. Without it, there would be no payloads in space.

Because the scope of the rocket propulsion field is far too complicated to be treated comprehensively in a single chapter, only a limited number of topics are covered here. This chapter presents the theory of rocket propulsion, rocket propellants, types of chemical rocket engines, and advanced propulsion techniques.

THEORY OF ROCKET PROPULSION

Newton's three laws of motion apply to all rocket-propelled vehicles. They apply to gas jets used for attitude control, to small rockets used for stage separation or for trajectory correction, and to large rockets used to launch a vehicle from the surface of the earth. They also apply to nuclear, electric, and other advanced types of rockets, as well as to chemical rockets. Newton's laws of motion are stated briefly as follows:

1. Bodies in uniform motion, or at rest, remain so unless acted upon by an external unbalanced force.
2. The force required to accelerate a body is proportional to the product of the mass of the body and the acceleration desired.
3. To every action there is an equal and opposite reaction.

These laws may be paraphrased and simplified in relating them to propulsion. For example, the *first law* says, in effect, that the engines must be adequate to overcome the inertia of the launch vehicle. The engines must be able to start

the vehicle moving and accelerate it to the desired velocity. Another way of expressing this for a vertical launch is to say that the engines must develop more pounds of thrust than the vehicle weighs. Some space missions require slowing the vehicle or changing its direction. An unbalanced force must be applied to accomplish these tasks. Newton's first law is universal and therefore applies not only to a vehicle at rest on a planet, but also to a vehicle in the so-called "weightless" condition of free flight.

Several forces must be considered when the *second law* is applied. For example, the accelerating force is the *net force* acting on the vehicle. This means that if a 100,000-lb vehicle is launched vertically from the earth with a 150,000-lb thrust engine, there is a net force at launch of 50,000 lb—the difference between engine thrust and vehicle weight. Here, the force of gravity is acting opposite to the direction of the thrust of the engine.

Propellants comprise approximately 90% of the vehicle's weight at launch. As the engines run, propellants are expended, decreasing the vehicle weight. Therefore, the net force acting on the vehicle increases, and the vehicle accelerates rapidly.

The acceleration and the resulting velocities attained during powered flight are shown in Figure 1. The acceleration and velocity are low at launch and just after launch due to the small net force acting at that time. But both acceleration and velocity increase rapidly as the propellants are ejected. When the first stage engine is shut off and staging occurs, acceleration drops sharply. When the second stage engine ignites, acceleration and velocity will again increase. As more propellants are ejected by the upper stage rocket engines, there will be rapid increases in acceleration and velocity. When the vehicle reaches the correct velocity (speed and direction) and altitude for the mission, thrust is terminated. Acceleration drops to zero after thrust termination, or burnout, and the vehicle begins free flight. For vehicles with three, four, or more stages, similar changes appear in both the acceleration and velocity each time staging occurs. Staging a vehicle increases the velocity in steps to the high values required for space missions.

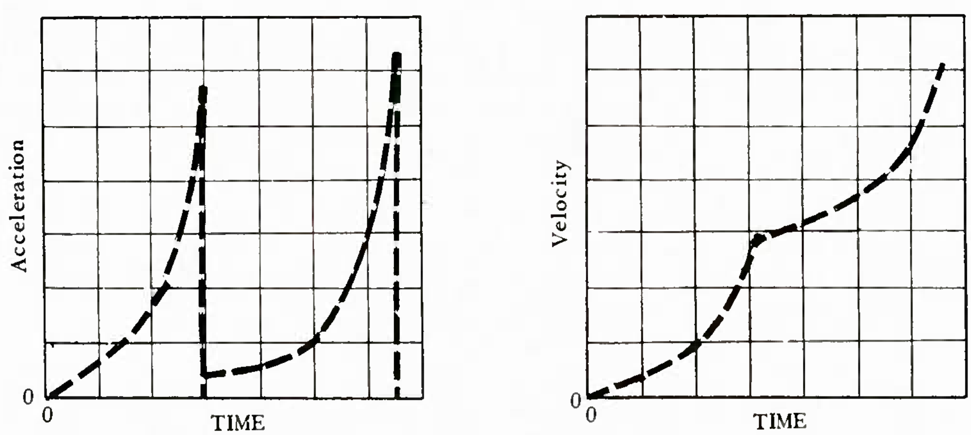


Figure 1. Powered flight of a typical 2-stage rocket.

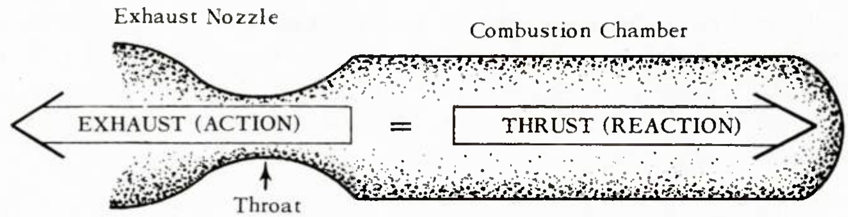


Figure 2. Action-reaction in a rocket motor.

Newton's second law also applies to a vehicle in orbit. Since the vehicle is in a "weightless" condition, the acceleration for a given thrust depends on the vehicle mass, because the vehicle still has mass even though it is "weightless." Remember that mass is the quantity of matter in a body, whereas weight is the force exerted on a given mass by a gravitational field. The mass of a body is the same everywhere, but its weight depends on its mass, the gravitational field, and its position in the gravitational field.

A given thrust will give a large acceleration to a vehicle with a small mass or a small acceleration to a vehicle with a large mass. Even a very small thrust (0.1 lb or less) operating for a long time can accelerate an earth-orbiting vehicle to the velocities needed to go to the other planets of our solar system. The section of this chapter on low thrust engines discusses such propulsion systems.

To relate Newton's *third,* or "action-reaction," *law* to rocket propulsion, consider what happens in the rocket motor (Fig. 2). All rockets develop thrust by expelling particles (mass) at high velocity from their exhaust nozzles. The effect of the ejected exhaust appears as a reaction force, called thrust, acting in a direction opposite to the direction of the exhaust.

Today, practically all exhaust nozzles, whether for a liquid or a solid-propellant rocket, use some form of the de Laval (converging-diverging) nozzle. It accelerates the exhaust products to supersonic velocities by converting some of the thermal energy of the hot gases into kinetic energy.

In the combustion chamber, the burning propellants have negligible unidirectional velocity but have high temperature and pressure. The high pressure forces the gases through the nozzle to the lower pressure outside the rocket. As the gases move through the *converging* section of the nozzle, their temperature and pressure decrease, and their velocity is increased to Mach 1 (the speed of sound) at the smallest cross-sectional area of the nozzle which is called the throat. The gases will attain sonic velocity at the nozzle throat if the combustion chamber pressure is approximately twice the throat pressure.

Since the speed of sound increases with an increase in the temperature of the propagating gas, both sonic velocity and actual linear gas velocity in the throat increase with an increase in gas temperature. Therefore, the higher the gas temperature at the throat, the higher the exhaust velocities that can be generated.

Thrust

Thrust (F) of a rocket is a sum of two terms, "momentum thrust" and "pressure thrust." Momentum thrust is the product of the *propellant rate of flow*

and the *velocity of the exhaust* relative to the rocket. Pressure thrust is the product of the maximum cross-sectional *area* of the divergent nozzle section and difference between *exit pressure* of the exhaust and the *ambient pressure* surrounding the rocket. The functional thrust equation below shows these relationships:

$$F = \underbrace{\frac{\dot{W}}{g} v_e}_{\text{Momentum Thrust}} + \underbrace{A_e (P_e - P_o)}_{\text{Pressure Thrust}} \qquad \text{(1) (Appendix E)}$$

In the above equation:

F = Thrust developed (lb)

$\dot{W}$ = Weight rate flow of propellants (lb/sec)
g = Acceleration of gravity at the earth's surface (32.2 ft/sec^2)
v_e = Velocity of gases at nozzle exit (ft/sec)
A_e = Cross-sectional area of nozzle exit (in^2)
P_e = Pressure of gases at nozzle exit (lb/in^2)
P_o = Ambient pressure (lb/in^2)

In high thrust rockets, which eject many pounds of propellant (as exhaust products) per second at velocities of several thousands of feet per second, the *momentum thrust is by far the dominant part* of the total thrust. It usually comprises more than 80% of the thrust being developed.

Nozzles and Expansion Ratio

The condition of "optimum expansion" occurs in a rocket nozzle when the pressure of the exhaust gases at the nozzle exit (P_e) is equal to the ambient (atmospheric) pressure (P_o). When $P_e = P_o$, the thrust output of the engine is the maximum that can be obtained at *that* altitude from *that* particular engine. For any given nozzle this condition can occur at only one altitude, i.e., where $P_e = P_o$. Therefore the altitude at which "optimum expansion" occurs depends upon the expansion ratio of the nozzle. "Expansion ratio" is defined as the area of the nozzle exit plane (A_e) divided by the area of the nozzle throat (A_t). It is designated by the letter epsilon. Thus:

$$\epsilon = \frac{A_e}{A_t}. \qquad (2)$$

Figure 3 shows the effect of nozzle expansion ratio on thrust. If the nozzle is cut off to the left of point B, the exit pressure is greater than ambient ($P_e > P_o$), and the nozzle is underexpanded (A_e/A_t is too small). The exhaust gases complete their expansion outside the nozzle.

If the nozzle is extended beyond point B (further increasing the expansion ratio), the exhaust gas exit pressure will be less than ambient ($P_e < P_o$), and the nozzle will be overexpanded. Net thrust will be less because the *pressure thrust loss,* due to the increase in A_e, *is greater than the momentum thrust gained* from the increase in v_e.

Thus, *for a given altitude* (P_o) and a fixed nozzle, maximum thrust is obtained when the nozzle has an expansion ratio so that $P_e = P_o$. As this same engine is flown to a higher altitude, the net thrust output will increase because of an increase in pressure thrust, but net thrust will not be as great as it *could* be if the nozzle expansion ratio could also be adjusted so that $P_e = P_o$ at all altitudes through which the engine operates.

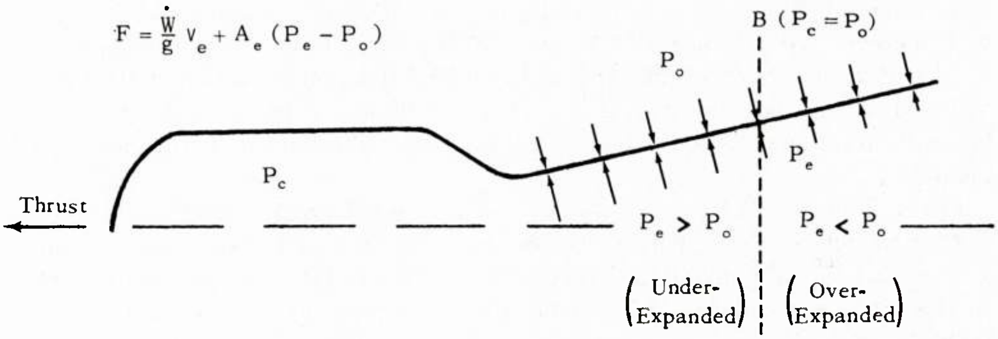


Figure 3. Nozzle expansion.

In summary, the altitude where $P_e = P_o$ is called the design altitude for a specific rocket engine. When $P_e = P_o$, the thrust output of the engine is the maximum that can be obtained at *that* altitude from *that* engine.

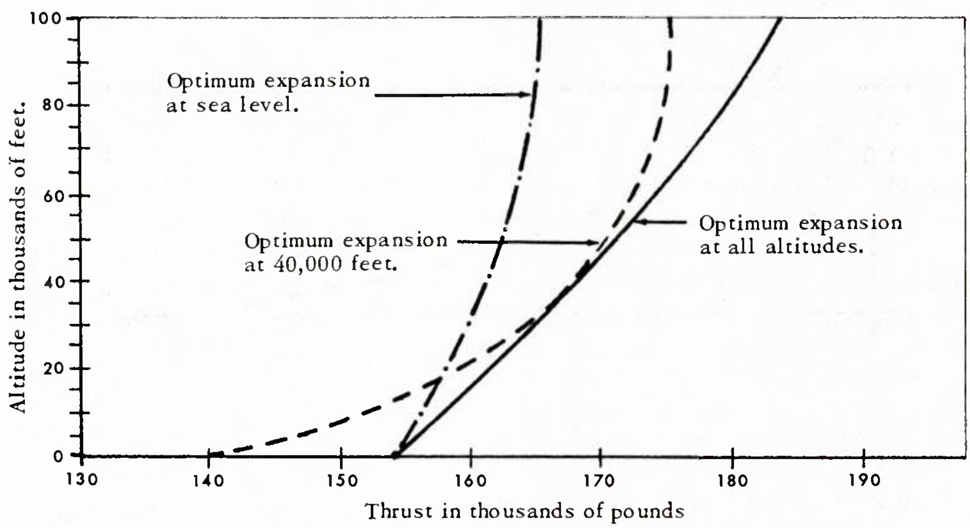


Figure 4. Thrust variation with altitude for nozzles of different expansion ratios.

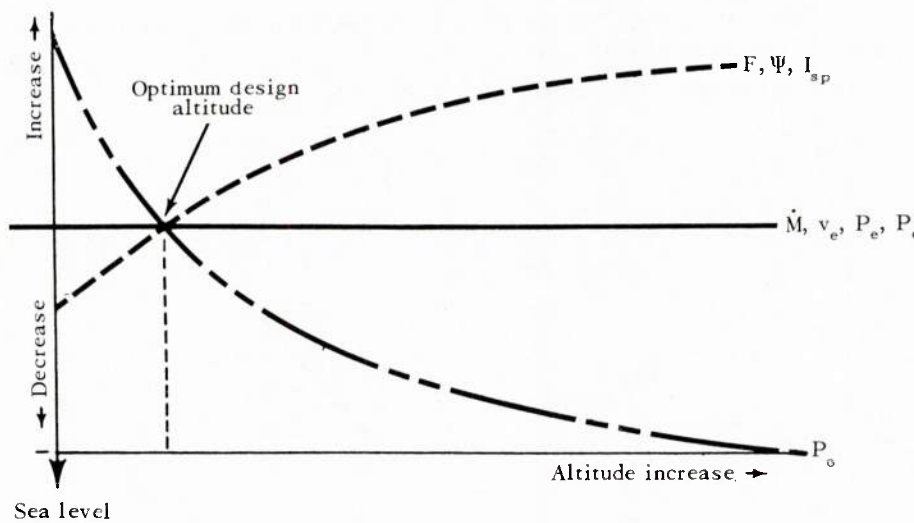
Nozzles can be designed for optimum expansion at sea level or at any higher altitude. The designer selects the altitude for optimum expansion that gives the best average performance over the powered flight portion of the vehicle trajectory. In

multistage rockets the expansion ratios vary in the vehicle stages which operate at different altitudes, with increasingly higher expansion ratios being used for those stages operating at higher altitudes.

Altitude Effects and Thrust Parameters

In evaluating the preceding information, remember that thrust with a given nozzle (fixed area ratio), *regardless* of the design altitude at which optimum expansion occurs, *increases* with altitude until P_o is essentially zero. Large booster engines use nozzles which are overexpanded at launch, achieve optimum expansion at 40,000 to 50,000 feet and have underexpansion above 50,000 feet. In this way they have higher thrust at more altitudes than if they were optimally expanded at launch. (See Figure 4.)

Figure 5 shows which parameters increase as a specific rocket climbs to altitude, which parameters tend to remain constant, and which parameters decrease in value. The rocket will reach an altitude where ambient pressure (P_o) becomes nearly zero. At this altitude, the thrust (F) and the specific impulse (I_{sp}) stabilize at a high value.



Legend

F		Thrust	v_e	=	Velocity of gases at nozzle exit
Ψ	=	Thrust-to-weight ratio	P_e	=	Pressure of gases at nozzle exit
I_{sp}	=	Specific impulse	P_c	=	Combustion chamber pressure
$\dot{M}$	=	Propellant mass rate of flow $= \frac{\dot{W}}{g}$	P_o	=	Ambient atmospheric pressure

Figure 5. A summary of the common rocket engine thrust parameters vs. altitude.

Specific Impulse

Thrust is also important in the definition of specific impulse:

$$I_{sp} = \frac{\text{Thrust (F)}}{\text{Weight rate flow of propellants } (\dot{W})} \qquad (3)$$

Thrust (F) is expressed in pounds, and propellant flow rate ($\dot{W}$) in pounds per second. Thus specific impulse (I_{sp}) is stated in seconds. If the propulsion system has an I_{sp} of 300 seconds, it produces 300 lbs of thrust for every pound of propellant burned per second. I_{sp} is one index of propulsive performance and is related to overall rocket performance.

Increasing I_{sp} improves the propulsion system's ability to increase vehicle velocity. This is the reason that it is frequently quoted to compare the performance of *similar propulsion systems*. I_{sp} should *not* be used alone to compare chemical and electrical or other dissimilar propulsion systems.

In designing a propulsion system, many compromises must be considered because of the many interrelated parameters. However, when a chemical propulsion system is designed only to improve I_{sp}, there are two basic approaches. One is to design a better rocket engine. The other is to use better propellants. These two methods are discussed later in this chapter.

Mass Ratio

Mass ratio is a structural design parameter. It is related to propulsion because the difference between initial (or launch) and final (or empty) vehicle weights is the weight of propellant expended.

$$\text{Mass ratio} = \frac{\text{Initial Weight}}{\text{Final Weight}} = \frac{\text{Weight at Engine Start}}{\text{Weight at Engine Shutdown}} = \frac{W_1}{W_2} \qquad (4)$$

For example, a single-stage, World War II rocket had a mass ratio of 3 to 1 (expressed as $\frac{3}{1}$) and a range of about 200 miles. Mass ratio for the rocket was computed using these figures:

Item	*Launch Weight (lb)*	*Empty Weight (lb)*
Payload	2,000	2,000
Propellants	16,000	0
Other dry weight	6,000	6,000
Total	24,000	8,000

$$\text{Mass ratio} = \frac{\text{Launch Weight}}{\text{Empty Weight}} = \frac{24{,}000}{8{,}000} = \frac{3}{1}$$

It appeared that the way to propel a payload to greater ranges was to build bigger and better rockets. A bigger rocket could carry more propellants, permit

longer thrust time, and achieve greater velocities and ranges. In improving the above rocket, the following ICBM design evolved:

Item	*Weight (lb)*	
Payload	200	(note decrease in payload)
Propellants	1,254,000	
Dry weight	29,800	
Launch Weight	1,284,000	

$$\text{Mass Ratio} = \frac{W_1}{W_2} = \frac{(1{,}284{,}000 \text{ lb})}{(30{,}000 \text{ lb})} = \frac{42.8}{1}$$

Obviously, mass ratio had to increase markedly to achieve ICBM ranges. High mass ratios have not been achieved with single stage rockets. A good single stage rocket may have a $\frac{10}{1}$ mass ratio. Mass ratios have been increased by using multistage rockets.

STAGING AND MASS RATIO.—The way that staging increases mass ratio is shown by the following ICBM based on design criteria similar to, but slightly better than, the World War II rocket.

	Item	*Weight (lb)*	
	Third stage		
Stage weight	Payload	200	= Vehicle weight at engine cut-off (W_2) (1,000 lb)
Stage weight	Dry weight	800	
Stage weight	Propellant	2,500	
	Total	3,500	= Vehicle weight at engine start (W_1)
	Second stage		
Stage weight	Payload	3,500	= Vehicle weight at engine cut-off (W_2) (10,000 lb)
Stage weight	Dry weight	6,500	
Stage weight	Propellant	25,000	
	Total	35,000	= Vehicle weight at engine start (W_1)
	First stage		
Stage weight	Payload	35,000	= Vehicle weight at engine cut-off (W_2) (65,000 lb)
Stage weight	Dry weight	30,000	
Stage weight	Propellant	162,500	
	Total	227,500	= Vehicle weight at engine start (W_1)

This rocket carried the 200-lb payload to ICBM ranges just as the scaled-up rocket previously described, but it had a gross weight at launch of only 227,500 lb.—about 17.7% of that of the scaled-up rocket. In the figures above, the initial weights (W_1) and final weights (W_2) for each stage are used to calculate the mass ratio for *each stage operation.* Remember that for each stage the initial weight is the weight of the *vehicle* when the stage's engines are started, and the final weight is the weight of the *vehicle* when that stage's engine is shut-off. There are three mass ratios to be considered:

The third mass ratio is for the final stage only:

$$\text{Third-stage mass ratio} = \frac{W_1}{W_2} = \frac{3{,}500 \text{ lb}}{1{,}000 \text{ lb}} = \frac{3.5}{1}$$

The second mass ratio is for the remainder of the vehicle for second-stage operation:

$$\text{Second-stage mass ratio} = \frac{35{,}000 \text{ lb}}{10{,}000 \text{ lb}} = \frac{3.5}{1}$$

The first mass ratio is for the whole vehicle for first-stage operation:

$$\text{First-stage mass ratio} = \frac{227{,}500 \text{ lb}}{65{,}000 \text{ lb}} = \frac{3.5}{1}$$

OVERALL MASS RATIO.—The overall mass ratio of a multistage rocket is the *product* of the stage mass ratios (See appendix E). Therefore, in the three-stage rocket the overall mass ratio is:

$$\text{Overall mass ratio} = \left(\frac{3.5}{1}\right)\left(\frac{3.5}{1}\right)\left(\frac{3.5}{1}\right) = \frac{42.8}{1}$$

Staging reduces the launch size and weight of the vehicle required for a specific mission and aids in achieving the high velocities necessary for ICBM and space missions. The velocity of the multistage vehicle at the end of powered flight is the sum of velocity increases produced by each of the various stages. The increases are added because the upper stages start with velocities imparted to them by the lower stages.

Thrust-to-Weight Ratio

Comparison of engine thrust to vehicle weight is expressed as a thrust-to-weight ratio (Ψ).

$$\Psi = \frac{\text{Thrust (F)}}{\text{Weight of vehicle (W)}} \tag{5}$$

A vehicle launched vertically cannot lift off the surface of the earth unless Ψ is greater than 1 ($F > W$). The larger the value of Ψ, the higher the initial vehicle acceleration.

$$a = (\Psi - 1) \text{ g's} \qquad \text{(6) (Appendix E)}$$

The Ψ of the Minuteman missile is approximately 2, and its initial acceleration is about 1 g, compared with a Titan II missile with a Ψ of about 1.4 and an initial acceleration of about 0.4g.

Mission Velocity Requirements

Figure 6 shows approximate values for the velocity and range relationships of a vehicle launched from the earth. Exact values depend on mission requirements, but a specific velocity is needed at the end of powered flight for each range—that

is, *I*ntermediate *R*ange *B*allistic *M*issile (IRBM) range—about 16,000 ft/sec; *I*ntercontinental *B*allistic *M*issile (ICBM)—24,000 ft/sec; and escape velocity near the earth's surface—36,700 ft/sec. Note that the velocity for ICBM range is very close to orbital velocities. In this area, small changes in velocity at the end of powered flight result in large changes in range.

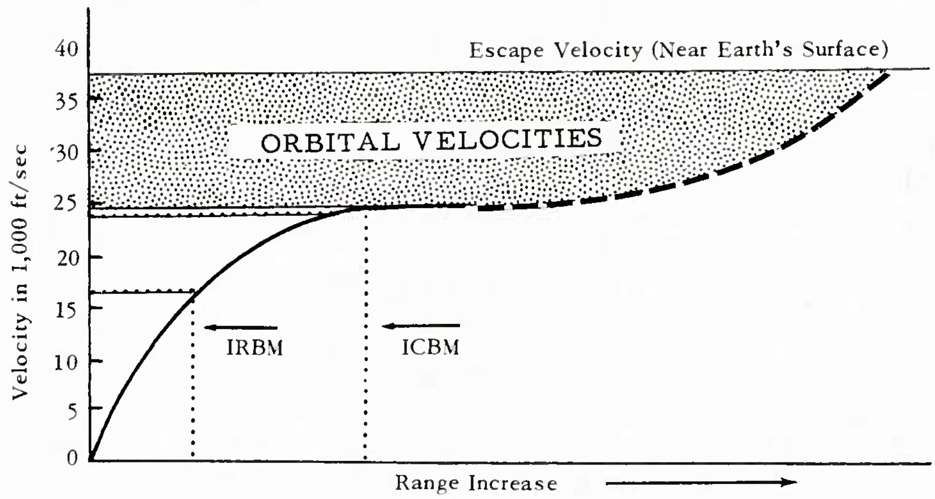


Figure 6. Velocity vs. range of a rocket.

IDEAL VEHICLE VELOCITY CHANGE

Specific impulse and mass ratio directly affect vehicle velocity. The vehicle performance equation shows how they are related to the magnitude of the ideal velocity change (Δv_i) of each stage at burnout or thrust termination.

$$\Delta v_i = I_{sp}\, g \ln \left(\frac{W_1}{W_2} \right) \qquad (7) \text{ (Appendix E)}$$

The ideal Δv is the velocity change the rocket would attain if there were *no gravity, no drag,* and if the *earth did not rotate.*

This simplified equation is derived from Newton's second law and is generally used to find approximate Δv's. The equation will not provide the exact Δv, since atmospheric drag, the earth's rotation, and the changing effects of gravity will cause variations. A precise solution is a computer calculation, but this equation gives a reasonable approximation. Due to losses, the actual Δv's of earth-launched vehicles will be 4,000 to 6,000 ft/sec lower than those computed with this equation. They will also vary with launch latitude and azimuth (see Appendix E).

Equation 7 shows Δv to be directly proportional to two factors: I_{sp} and the natural logarithm (ln) of the mass ratio (MR). In other words, higher Δv's can be achieved by increasing either I_{sp} or MR, or both. The equation does not consider physical size of the rocket or propulsion system. Therefore, in theory, different size rockets with identical I_{sp} and MR would achieve the same ideal Δv. Obviously,

larger payloads require larger vehicles and higher thrust engines to achieve the same range or velocity.

The "g" in the equation comes from the conversion of mass to weight at the surface of the earth. Calculations in space use an altitude I_{sp}, which includes the increase in thrust with altitude but measures W at the surface of the earth. Therefore, g is *always* 32.2 ft/sec^2 in propulsion calculations.

Equation 7 applies to all rocket vehicles. For a one-stage rocket the Δv would be the final velocity at thrust termination with the mass ratio based on launch and burnout conditions. In multistage rockets, the ideal Δv is the sum of the Δv's developed by the various stages, with the I_{sp} and MR of each stage used to compute the Δv for that stage's burning time.

The equation may also be used to compute Δv for all space vehicles, including those with a stop and restart capability. When power is applied in flight, a mass ratio combining initial and final weights of the space vehicle for each powered phase is used.

ACTUAL VEHICLE VELOCITY CHANGE

Actual launch vehicle velocity can be expressed as:

$$\Delta v_a = \Delta v_i - \Delta v_l + v_r \qquad (8)$$

In this equation:

Δv_a = Actual Δv

Δv_i = Ideal Δv (from equation 7)

Δv_l = Loss in Δv (gravity and drag) (See Appendix E)

v_r = Variation in Δv due to earth's rotation (See Appendix E)

Two types of problems that may be solved with equation 7 are presented on succeeding pages. A slide rule, math tables, or the graph given in Figure 7 can be used to determine natural logarithms. It should be noted that, since the graph is actually a plot of natural logarithm (1n) values, it may be extended to accommodate values beyond the limits shown. The graph is plotted so that mass ratios are on the vertical axis and the natural logarithms are on the horizontal axis. For example, if $(W_1/W_2) = 8$, this is found on the vertical axis. Then the 1n of 8 = 2.08 is found on the horizontal axis.

Sample Rocket Performance Calculation

A rocket was launched from the earth with an initial weight (W_1) of 296,000 lb. At the end of powered flight, its final weight (W_2) was 40,000 lb. Assuming no earth rotation and no velocity loss due to drag or gravity, calculate the magnitude of the ideal velocity (or Δv_i) at the end of powered flight.

$$I_{sp} = 300 \text{ sec} \qquad \text{and} \qquad g = 32.2 \text{ ft/sec}^2$$

Use the equation: $\Delta v_i = I_{sp}\, g \, 1n \left(\frac{W_1}{W_2}\right)$

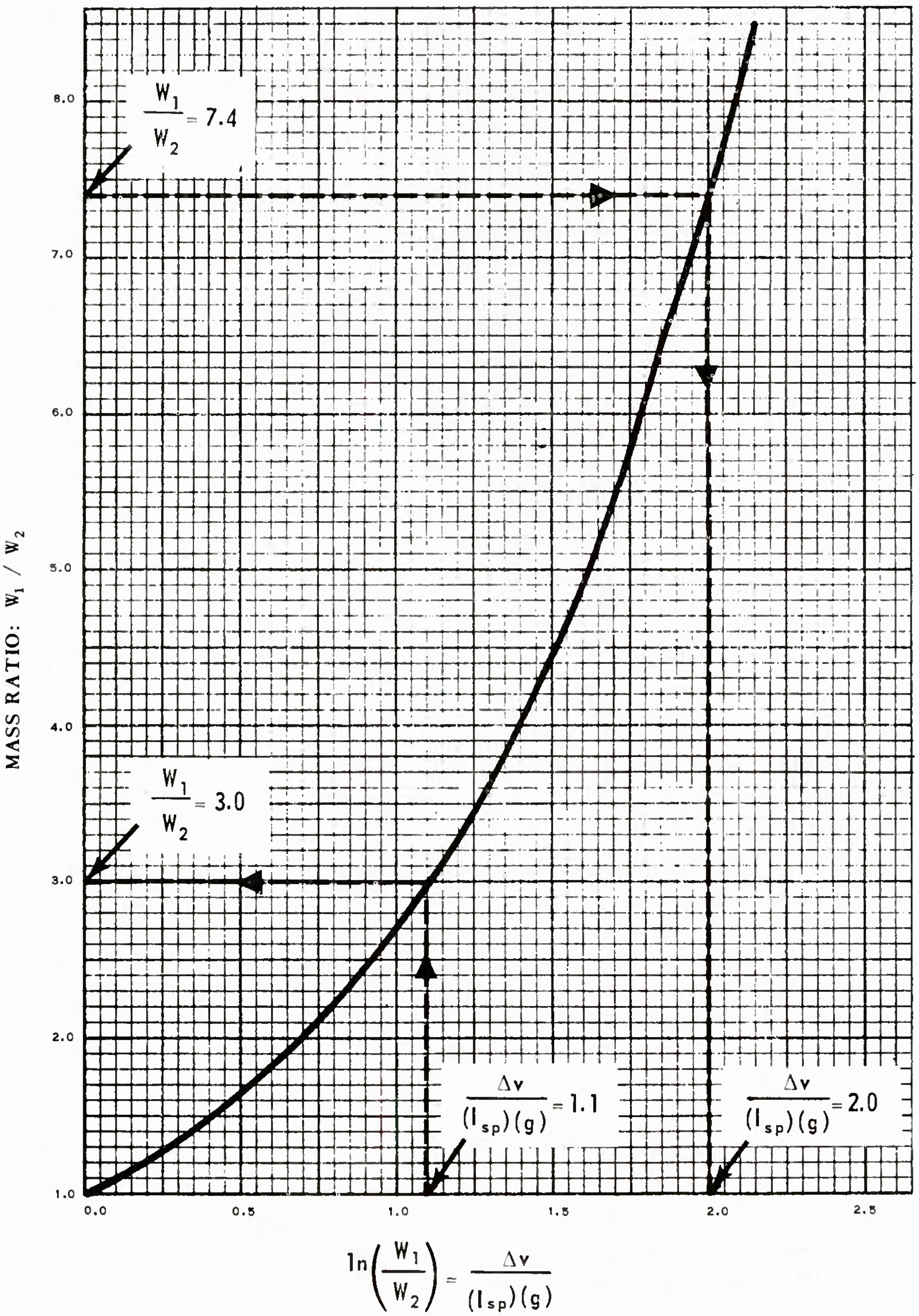


Figure 7. Mass ratio vs. $\Delta v/(I_{sp})$ (g) or ln (W_1/W_2).

Slide-rule method

a. Substitute

$$\Delta v_i = (300)(32.2) \ln \frac{296{,}000}{40{,}000}$$

$$\Delta v_i = (300)(32.2) \ln 7.4$$

b. Find: ln 7.4 on slide rule. (above 7.4 on the LL3 scale read 2.0 on the "D" scale.)

$$\ln 7.4 = 2.0$$

c. Then:

$$\Delta v_i = (300)(32.2)(2.0)$$

$$\Delta v_i = 19{,}320 \text{ ft/sec}$$

Graph method

a. Transpose equation to read:

$$\frac{\Delta v_i}{(I_{sp})(g)} = \ln\left(\frac{W_1}{W_2}\right)$$

b. Calculate value of W_1/W_2

$$\frac{296{,}000}{40{,}000} = 7.4$$

c. Use chart, Figure 7.

d. Enter the graph where $W_1/W_2 = 7.4$ and then

e. Read the value of:

$$\frac{\Delta v_i}{(I_{sp})(g)} = 2.0$$

f. Then, solving for Δv_i

$$\Delta v_i = (I_{sp})(g)\ (2.0)$$

$$\Delta v_i = (300)(32.2)(2.0)$$

$$\Delta v_i = 19{,}320 \text{ ft/sec}$$

Sample problem for Changing an Orbit

(Finding propellants required)

A space vehicle is coasting in orbit. It has a restart capability and 3,500 lb of propellants on board. A change in orbit is desired. Calculate the propellant required for the orbit change and whether the maneuver can be completed, based on the following data.

Δv_a = 14,170 ft/sec = Magnitude of velocity change required to achieve the new orbit

I_{sp} = 400 sec = Specific impulse of the vehicle's engine at *altitude.*

W_1 = 5,000 lb = Current weight of the vehicle in orbit (measured at surface of the earth)

$g = 32.2 \dfrac{\text{ft}}{\text{sec}^2}$ = Acceleration due to gravity at earth's surface

1. Use the equation: $\Delta v_a = I_{sp}\, g \ln (W_1/W_2)$ (There are no drag or gravity losses in space so $\Delta v_i = \Delta v_a$)

2. Transpose to read: $\dfrac{\Delta v}{(I_{sp})(g)} = \ln\left(\dfrac{W_1}{W_2}\right)$

3. Solve for value of: $\Delta v/(I_{sp})(g)$

$$\frac{\Delta v}{(I_{sp})(g)} = \frac{14{,}170}{(400)(32.2)} = 1.1; \text{ therefore } 1.1 = \ln(W_1/W_2)$$

4. Complete solution by either of following methods

Slide rule method	*Graph method*
a. Since: $\ln(W_1/W_2) = 1.1$ (below 1.1 on "D" scale read 3.0 on the LL3 scale.)	a. Use graph, Figure 7.
b. Then: $W_1/W_2 = 3.0$	b. Enter the graph where: $\Delta v/(I_{sp})(g) = 1.1$ and then Read the value of: W_1/W_2, which is 3.0
c. Then solving for W_2:	c. Then solving for W_2:
$W_2 = \frac{W_1}{3.0} = \frac{5{,}000}{3.0} = 1{,}667$ lbs	$W_2 = \frac{W_1}{3.0} = \frac{5{,}000}{3.0} = 1{,}667$ lb

d. For both above methods, let W_p = lb of propellants consumed;

e. Then: $W_p = W_1 - W_2 = 5{,}000 - 1{,}667 = 3{,}333$ lb of propellants consumed.

The orbital change can be made because the amount of propellants required is less than the 3,500 lb which are available.

PARKING ORBITS

Most manned missions which proceed to orbit or go farther into space begin by placing a payload into a parking orbit about the earth. The time in the parking orbit is used to wait for appropriate phase angles, for adjustments in launch windows, or for verifying equipment performance before proceeding on the next leg of the mission.

Somewhere between 85 and 100 NM is considered optimum for a parking orbit from a drag and gravity standpoint. The choice of 100 NM allows optimum time in orbit for a manned capsule.

Although the prediction of orbital lifetime is as yet an inexact sciencc, these approximate values illustrate thc choices available for parking orbits:

Altitude (NM)	*Expected Time in Orbit (Days)*
85	½
100	3
150	35
200	200
300	4,000

ROCKET PROPELLANTS

Propellants are the working substances that rocket engines use to produce thrust. These substances may be liquid, solid, or gas, but liquids or solids are usually used because they permit more chemical energy to be carried in a particular rocket. Large bulky tanks would be required to contain only a small mass of compressed gas.

This section of the text considers only working substances which are accelerated by the energy released in the chemical combustion (burning) of these substances.

The process involves a fuel and oxidizer reacting chemically to produce high temperature, high pressure gases. Such fuels and oxidizers are called chemical propellants.

Two chemical propellant subject areas are important: theoretical performance characteristics and energy content.

Theoretical Performance of Chemical Propellants

Consider a simplified picture of what happens in a rocket combustion chamber and nozzle (Fig. 8). Burning the propellants releases large amounts of energy and produces high temperature, high pressure gases. The high temperature at point A in the combustion chamber is associated with very rapid motion of the gas molecules. These molecules have high speeds but move in random directions. As they approach the nozzle throat (B), their motion is less random, and they move toward the nozzle exit. At the nozzle exit (C), the largest velocity component of the molecules is parallel to the nozzle axis. The combustion chamber converts the chemical energy of the propellants to high temperature random motion of the gas molecules, and the nozzle orients the velocity or kinetic energy (energy of motion) which gives the rocket its thrust.

As pointed out before, the velocity change of a space vehicle is a function of specific impulse (I_{sp}) and mass ratio. A higher I_{sp} increases the magnitude of the vehicle Δv for a given mass ratio or decreases the mass ratio necessary for a

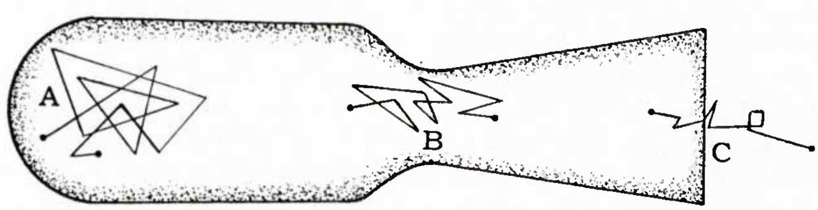


Figure 8. Velocity of gas molecules after combustion in a rocket engine.

given vehicle Δv. Thus, I_{sp} is a measure of how well an engine converts chemical energy into velocity.

Since each mission in space is associated with a required Δv, I_{sp} determines if it is possible to perform a space mission with the mass ratios which are obtainable. For example, a mission to launch from the earth, land on the moon, and return to earth requires a series of Δv's totaling about 59,000 mph. If the engines use propellants with low I_{sp} and can produce Δv's of only 55,000 mph for a particular mass ratio or payload, the space vehicle cannot complete the mission. Engines and propellants with higher I_{sp} are needed for more difficult space missions.

Theoretical Specific Impulse

The theoretical specific impulse of chemical propellants in an ideal rocket is as follows:

$$I_{sp} = 9.797 \sqrt{\left(\frac{k}{k\text{-}1}\right)\left(\frac{T_c}{m}\right)\left[1-\left(\frac{P_e}{P_c}\right)^{\frac{k-1}{k}}\right]} \qquad (9) \text{ (Appendix E)}$$

Where:

I_{sp} = Specific impulse (sec)
k = Average ratio of specific heats (C_p/C_v) of the combustion products. C_p is the specific heat of gases at constant pressure. C_v is the specific heat of gases at constant volume.
P_e = Nozzle-exit pressure (psia)
P_c = Combustion-chamber pressure (psia)
m = Average molecular weight of combustion products (lb/mole)
T_c = Combustion chamber temperature (°R). (Degrees Rankine is the absolute temperature and is equal to temperature in degrees F + 459.7°.)

The actual values of C_p, C_v, and k depend on the composition of the gas and its temperature. The combustion products in a rocket chamber and nozzle are a mixture of gases which vary in temperature from about 5,500°R in the combustion chamber to about 3,000°R at the nozzle exit; k is an average value of C_p/C_v for these temperatures (1.2 - 1.33 for chemical engines).

A molecule is the smallest quantity of matter which can exist by itself and retain all the properties of the original substance. For example, a molecule of water is the smallest quantity which has all the properties of water. A molecule of water contains two atoms of hydrogen whose combined atomic weight is about 2, and one atom of oxygen whose atomic weight is 16. Since a molecule is such a small amount of a substance, a mass numerically equal to the combined atomic weights, the molecular weight (called the mole) is used to give the equation a workable mass. In the English system of units a mole of water would be 18 pounds, and is usually called a pound-mole of water. Since the combustion products in a chemical rocket are a mixture of many gases, such as water vapor, carbon monoxide, carbon dioxide, hydrogen, and oxygen, *the average molecular weight of the combustion products* is used in the equation for theoretical specific impulse.

The combustion chamber temperature is the temperature obtained from the reaction of the oxidizer and fuel. This temperature depends on the mixture ratio which is computed as follows:

$$\text{Mixture ratio (r)} = \frac{\text{Weight flow rate of oxidizer } (\dot{W}_o)}{\text{Weight flow rate of fuel } (\dot{W}_f)} \qquad (10)$$

There are high and low limits to the possible mixture ratios for an oxidizer and fuel combination. If there is not enough oxidizer, or if there is too much oxidizer, combustion does not take place. A mixture ratio which produces complete combustion of both oxidizer and fuel gives the highest temperature. Too much oxidizer,

even within the combustion mixture ratio limits, results in a lower temperature and excess oxidizer in the combustion products. Too little oxidizer results in a lower temperature and unburned fuel in the combustion products. Thus, mixture ratio, combustion chamber temperature, and average molecular weight of the combustion products are all interrelated.

Consider the relative effects of the ratio of specific heats, pressure, average molecular weight of combustion products, and combustion chamber temperature on theoretical I_{sp} for these conditions:

$k = 1.2$ $\quad m = 20$ lb/mole $\quad P_c = 1{,}000$ psia

$T_c = 6{,}000°R.$ $\quad P_e = 14.7$ psia

$$I_{sp} = 9.797 \sqrt{\left(\frac{1.2}{0.2}\right)\left(\frac{6{,}000}{20}\right)\left[1 - \left(\frac{14.7}{1{,}000}\right)^{\frac{0.2}{1.2}}\right]}$$

$$I_{sp} = 9.797\sqrt{(6)\ (300)\ (0.504)}$$

The ratio T_c/m (300) dominates the equation and:

$$I_{sp} \alpha \sqrt{\frac{T_c}{m}} \qquad (11)$$

This states that the theoretical I_{sp} of chemical propellants is directly proportional to the square root of the ratio of the combustion chamber temperature and the average molecular weight of the combustion products. The actual calculation of theoretical I_{sp} is complicated. The solution requires an extensive computation which is best done with an electronic computer.

The relationship among specific impulse, combustion chamber temperature, and the molecular weight of the combustion products is shown in Figure 9. Note that I_{sp} increases as the value of T_c/m increases.

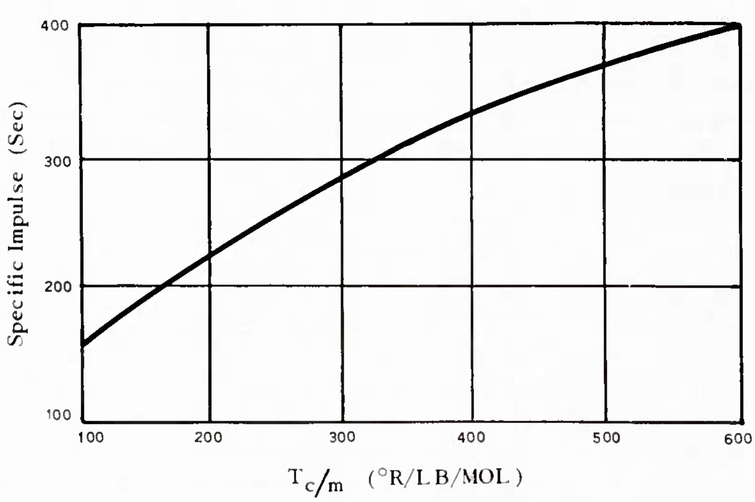


Figure 9. Specific impulse vs. T_c/M.

The highest T_c may not produce the highest I_{sp}, since a corresponding increase in the average molecular weight of the combustion products may result in a smaller value for T_c/m. For example, most rockets using a hydrocarbon fuel like RP-1 (highly refined kerosene) burn a fuel-rich mixture which does not produce the highest T_c. This fuel-rich mixture produces low molecular weight gases like carbon monoxide and results in low average molecular weight combustion products, a high value for T_c/m, and a high I_{sp}. Burning a mixture of liquid oxygen (LOX) and RP-1 which produces the highest T_c results in: high molecular weight gases like carbon dioxide; a lower value for T_c/m; and a lower I_{sp}.

The importance of low molecular weight combustion products effectively limits rocket fuels to chemical compounds of light elements, such as hydrogen, lithium, boron, carbon, and nitrogen. The use of elements heavier than aluminum (atomic weight 27) generally results in lower I_{sp}. Since hydrogen has the lowest molecular weight of all the elements, a propellant with a high hydrogen content is desirable.

In actual practice, the highest performance rocket fuels are relatively rich in hydrogen. Table 1 lists the *theoretical* I_{sp}'s for various typical liquid propellant combinations. These are calculated for the following conditions: a combustion chamber pressure of 1000 psia, an optimum nozzle expansion ratio, an ambient pressure of 14.7 psia and a state of shifting equilibrium. The state of shifting equilibrium assumes a condition of changing chemical composition of gaseous products throughout expansion in the rocket nozzle. This changing chemical composition results in a higher I_{sp} than exists when the chemical composition of the gaseous products is fixed throughout expansion in the nozzle. The latter condition is called frozen equilibrium. The actual composition is somewhere between these two conditions.

Density Impulse

In some instances, a number of different propellant combinations may provide the same I_{sp}. Choosing a combination may then depend upon a parameter called density impulse. Density impulse is the product of specific impulse and specific gravity (SG) of the propellant:

$$I_d = (I_{sp})\,(SG) \tag{12}$$

This parameter is used to relate propulsion system performance to the volume of the tank required to contain the propellants. Given a choice among propellant combinations each having the same I_{sp}, the combination with the highest SG requires smaller propellant tanks.

Total Impulse

Total impulse is directly related to the vehicle's ideal velocity change. Total impulse (I_t) relates the propulsion systems' thrust (F) to the time of operation or burning time (t_b) and is defined as:

$$I_t = (F)\,(t_b) \tag{13}$$

Note that interchanging the values of thrust and operating time does not change the total impulse. For vertical flight the thrust must exceed the weight of the vehicle in order for it to accelerate. But in orbit, either a high or low thrust would accelerate the vehicle to the same final velocity, provided the total impulse were the same. A higher thrust system accomplishes this through higher acceleration but for a shorter period of time than a low thrust system.

Total impulse can also be defined as:

$$I_t = (I_{sp})\ (W_p) \tag{14}$$

From this it is possible to see that this parameter is used to relate propulsion system performance to the allowable weight for propellants. Given a choice among propellant combinations each having the same I_{sp}, the combination having the largest weight would produce the greatest change in velocity.

Characteristics and Performance of Propellants

Rocket engines can operate on common fuels such as gasoline, alcohol, kerosene, asphalt or synthetic rubber, plus a suitable oxidizer. Engine designers consider fuel and oxidizer combinations having the energy release and the physical and handling properties needed for desired performance. Selecting propellants for a given mission requires a complete analysis of: mission; propellant performance, density, storability, toxicity, corrosiveness, availability and cost; size and structural weight of the vehicle; and payload weight.

LIQUID PROPELLANTS.—The term "liquid propellant" refers to any of the liquid working fluids used in a rocket engine. Normally, they are an oxidizer and a fuel but may also include catalysts or additives that improve burning or thrust. Generally, liquid propellants permit longer burning time than solid propellants. In some cases, they permit intermittent operation; that is, combustion can be stopped and started by controlling propellant flow.

Many liquid combinations have been tested; but no combination has all these desirable characteristics:

1. Large availability of raw materials and ease of manufacture.
2. High heat of combustion per unit of propellant mixture. (For high T_c).
3. Low freezing point (Wide range of operation).
4. High density (Smaller tanks).
5. Low toxicity and corrosiveness (Easier handling and storage).
6. Low vapor pressure, good chemical stability (Simplified storage).

Liquid propellants are classified as monopropellants, bipropellants, or tripropellants.

A *monopropellant* contains a fuel and oxidizer combined in one substance. It may be a single chemical compound, such as nitromethane, or a mixture of several chemical compounds, such as hydrogen peroxide and alcohol. The compounds are stable at ordinary temperatures and pressures, but decompose when

heated and pressurized, or when the reaction is started by a catalyst. Monopropellant rockets are simple since they need only one propellant tank and the associated equipment.

A *bipropellant* is a combination of fuel and oxidizer which are not mixed until after they have been injected into the combustion chamber. At present, most liquid rockets use bipropellants. In addition to a fuel and oxidizer, a liquid bipropellant may include a catalyst to increase the speed of the reaction, or other additives to improve the physical, handling, or storage properties. Some bipropellants use a fuel and an oxidizer which do not require an external source of ignition but ignite on contact with each other. These propellants are called *hypergolic.*

A *tripropellant* has three compounds. The third compound improves the specific impulse of the basic bipropellant by increasing the ratio T_c/m.

Liquid propellants are also commonly classified as either cryogenic or storable propellants.

A *cryogenic propellant* is one that has a very low boiling point and must be kept very cold. For example, liquid oxygen boils at $-297°$ F, liquid fluorine at $-306°$ F, and liquid hydrogen at $-423°$ F. These propellants are loaded into a rocket as near launch time as possible to reduce losses from vaporization and to minimize problems caused by their low temperatures.

A *storable propellant* is one which is liquid at normal temperatures and pressures and which may be left in a rocket for days, months, or even years. For example, nitrogen tetroxide (N_2O_4) boils at 70° F., unsymmetrical dimethylhydrazine (UDMH) at 146° F., and hydrazine (N_2H_4) at 236° F. However, the term "storable" means storing propellants on earth. It does not consider the problems of storage in space.

One of the cryogenic propellants, LOX, is used with RP–1 in many rocket engines. The H–1 and F–1 engines of the NASA Saturn vehicles use this combination.

Liquid hydrogen (LH_2) and LOX comprise a cryogenic bipropellant. It is used in upper stage engines, such as the RL–10 (Centaur engine), and in the J–2 of the Saturn V.

The 50% UDMH-50% hydrazine fuel (Aerozine 50) with nitrogen tetroxide (N_2O_4) oxidizer is the storable hypergolic bipropellant used in the Titan III. It is classed as a bipropellant since the fuel contains two compounds to improve handling properties rather than to improve I_{sp}.

The fluorine (LF_2) and LH_2 bipropellant with an I_{sp} of 410 seconds (Table 1) shows the improved performance of cryogenics. As a group, they have higher I_{sp} than the storable propellants.

The I_{sp} values in Table 1 represent the maximum *theoretical* values for normal test conditions, which include engine operation at sea level. Actual engines using these propellants at sea level achieve 85 to 92 percent of these values. Engines operating near design altitude frequently achieve specific impulses which exceed these values.

For example, one version of the Atlas engine using RP–1 and LOX is designed for optimum expansion at 100,000 ft. altitude. The I_{sp} of this engine is 215

TABLE 1

Specific Impulse of Liquid Propellant Combinations
(Units are given in seconds)

Oxidizer	FUEL					
	Ammonia	RP-1	UDMH	50% UDMH and 50% hydrazine	Hydrazine (N_2H_4)	Hydrogen*
Liquid oxygen*	294	300	310	312	313	391
Chlorine trifluoride	275	258	280	287	294	318
95% hydrogen peroxide and 5% water	262	273	278	279	282	314
Red fuming nitric acid (15% NO_2)	260	268	276	278	283	326
Nitrogen tetroxide	269	276	285	288	292	341
Fluorine*	357	326	343		363	410

* cryogenic

TABLE 2

Comparison of Payloads for a Three-Stage Launch Vehicle Using Conventional Propellants With One Using High-Energy Propellants in the Upper Stages (Gross weight about one million pounds)

MISSION	PAYLOAD IN TONS	
	Conventional Upper Stages	*High-Energy Upper Stages*
Low earth orbit		
1. Ideal velocity ≈ 30,000 ft/sec	15	32
2. Ideal velocity ≈ 34,000 ft/sec	9	21
Escape		
3. Ideal velocity ≈ 41,000 ft/sec	3	9

sec at sea level and 309 sec at 80,000 ft, compared with the 300 sec listed in Table 1. I_{sp} reported for rocket engines, especially if much higher than the values in Table 1, must be analyzed to determine the altitude and other conditions for which the values are tablulated.

Several of the liquid propellants have theoretical I_{sp} approximately one-third higher than the conventional LOX and RP–1. LOX with LH_2, and LF_2 with LH_2, are examples of bipropellants called *high-energy* propellants. The term "high-energy" evolved from efforts to develop high-performance propellants. All the upper stages of large launch vehicles like the Saturn IB and Saturn V use these propellants.

High-energy propellants in the upper stages of large rockets increase the payload or the mission capability of the vehicle. Consider two three-stage launch vehicles with the same initial gross weight (Table 2). One has high-energy pro-

pellants in the upper stages; the other has conventional LOX and RP–1 in the upper stages. The one using high-energy propellants can carry a heavier payload for a given mission, or can perform a more difficult mission with the same payload. A three-stage vehicle with a gross weight of about one million pounds will have the theoretical payloads shown. For mission number 1, the vehicle using conventional propellants has less than half the payload of the one with high-energy propellants. For mission number 3, the conventional vehicle has only one-third the payload of the high-energy vehicle. Also, the conventional vehicle has the same payload (9 tons) for mission number 2 that the high-energy one has for mission number 3. Increasing the number of stages of the vehicle with LOX/RP–1 will increase the payload for the same initial gross weight, but will not approach that of the high energy propellant vehicle.

High-energy fuel tanks are designed differently from those for conventional fuel. Hydrogen tanks are large and bulky because hydrogen has low density and low molecular weight. Therefore, upper stage weight and volume is larger if the upper stage uses hydrogen than if it uses conventional propellants.

Although high-energy propellants in upper stages increase mission capability, these propellants have high reactivity, thermal instability, and low temperatures. Because fluorine is very corrosive and toxic, it is difficult to handle and store. Fluorine, hydrogen, and oxygen are liquids only at very low temperatures. These low temperatures cause many metals to lose their strength, and may cause the freezing of handling equipment such as valves. Fluorine is found in relatively large quantities in nature but is expensive to concentrate in a free state. Because of these and other problems, high-energy propellants are usually used only in upper stages. They are expensive as first stage propellants, and are difficult to store in space vehicles.

SOLID PROPELLANTS.—Solid propellants burn on their exposed surfaces to produce hot gases. Solids contain all the substances needed to sustain combustion. Basically, they consist of either fuel and oxidizer which do not react below some minimum temperature, or of compounds that combine fuel and oxidizer qualities (nitrocellulose or nitroglycerin). These materials are mixed to produce a solid with the desired chemical and physical characteristics.

Solid propellants are commonly divided into two classes: composite (or heterogeneous), and homogeneous.

Composites are heterogeneous mixtures of oxidizer and organic fuel binder. Small particles of oxidizer are dispersed throughout the fuel. The fuel is called a binder because the oxidizer has no mechanical strength. Neither fuel nor oxidizer burns well in the absence of the other. Usually a crystalline, finely ground oxidizer such as ammonium perchlorate is dispersed in an organic fuel such as asphalt; the oxidizer is about 70 to 80 percent of the total propellant weight. There are a large number of propellants of this type.

Homogeneous propellants have oxidizer and the fuel in a single molecule. Most are based on a mixture of nitroglycerine and nitrocellulose, and are called "double-base propellants." The term distinguishes these propellants from many gunpowders which are based on either one or the other of the components. Nitroglycerine is too sensitive to shock and has too much energy to be used safely by itself in an engine. However, it forms a suitable propellant when combined

with the less energetic but more stable nitrocellulose. The major components of a typical double-basc propellant are:

Component	*Percent of total*
Nitrocellulose	51.38% (propellant)
Nitroglycerine	43.38% (propellant)
Diethyl phthalate	3.09% (plasticizer)
Potassium nitrate	1.45% (flash depressor)
Diphenylamine	0.07% (stabilizer)
Nigrosine dye	0.10% (opacifier)

Note the additives which control physical and chemical properties. Each additive performs a specific function. The *plasticizer* improves the propellant's structural properties. The *flash depressor* cools the exhaust gases before they escape to the atmosphere and promotes smooth burning at low temperatures. The *stabilizer* absorbs the gaseous products of slow decomposition and reduces the tendency of the propellant to absorb moisture during storage. The *opacifier* prevents heat transfer by radiation to sections of the propellant which have not started to burn. (Small flaws in the propellant can absorb enough heat through radiation to ignite the propellant internally, producing enough gas to break it up if an opacifier is not present.)

An ideal solid propellant would possess these characteristics:

1. High release of chemical energy.
2. Low molecular weight combustion products.
3. High density.
4. Readily manufactured from easily obtainable substances by simple processes.
5. Safe and easy to handle.
6. Insensitive to shock and tempcrature changes and no chemical or physical deterioration while in storage.
7. Ability to ignite and burn uniformly over a wide range of operating temperatures.
8. Nonhygroscopic (non-absorbent of moisture).
9. Smokeless and flashless.

It is improbable that any propellant will have all of these characteristics. Propellants used today possess some of these characteristics at the expense of others, depending upon the application and the desired performance.

The finished propellant is a single mass called a *grain* or stick. A solid propellant rocket has one or more grains which constitute a charge in the same chamber. The use of solid propellants was limited until the development of high-energy propellants, and of the processing techniques for making large grains. Now, single grains are made in sizes up to 22 ft. in diameter.

In addition to being composite or homogeneous, solid propellants are also classed as restricted or unrestricted. They are *restricted* burning charges when inhibitors are used to restrict burning on some surfaces of the propellant. Inhibitors are chemicals which do not burn or which burn very slowly. Controlling the burning area in this manner lengthens the burning time and results in lower thrust. An inhibitor applied to the wall of the combustion chamber reduces heat transfer to the wall and is called a liner.

Charges without an inhibitor are *unrestricted* burning charges. These burn on all exposed surfaces simultaneously. The unrestricted grain delivers a large thrust for a short time, whereas a restricted grain delivers smaller thrust for a longer time. Today, most large solid propellant rockets contain restricted burning charges.

The operating pressure, thrust, and burning time of a solid propellant rocket depend upon: the chemical composition of the propellant; its initial grain temperature; the gas velocity next to the burning surface; and the size, burning surface, and geometrical shape of the grain. A given propellant can be cast into different grain shapes with different burning characteristics.

The thrust of a rocket is proportional to the product of the exhaust velocity and the propellant flow rate. Large thrust requires a large flow rate which is produced by a large burning surface, a fast burning rate, or both. The burning rate of a solid propellant is determined by the speed of the flame passing through it in a direction perpendicular to the burning surface. Burning rate depends on the initial grain temperature, and upon the operating chamber pressure. A solid rocket operates at a higher chamber pressure and thrust when the propellant is hot than when it is cold, but it will burn for a shorter time. The converse is also true. If the chamber pressure is below a minimum value, the propellant will not burn.

Variation in the geometric shape of the grain changes the burning area and thrust output. Figure 10 shows several typical grain shapes.

Basically, there are three ways in which the burning area can change with time. If the area increases as burning progresses, the thrust increases with time, and the grain is called a *progressive* burning grain. If the burning area decreases with time, thrust decreases, and the grain is called a *regressive* burning grain. If the area remains approximately constant during burning, thrust is constant, and the grain is called a *neutral* burning grain. In each case, the type of burning determines how the thrust level changes with time, following initial thrust buildup.

Regressive burning does not produce as much peak acceleration as does neutral or progressive burning, because the thrust decreases as vehicle weight decreases. Typical progressive, neutral, and regressive grains are shown in Figure 11. Lower accelerations are required when rockets are used to propel payloads which cannot withstand high acceleration loading.

The *progressive* grain is a bored cylinder inhibited on the ends and at the chamber wall. All burning occurs on the surface of the central port area. Both the burning area and thrust increase as the propellant burns.

The *neutral* grain is a rod and tube inhibited on the ends and at the chamber wall. All burning occurs on the outside of the rod and on the inside of the tube. As burning progresses, the increasing area on the tube balances the decreasing area on the rod so that total area and thrust remain constant.

The *regressive* grain is a cylinder inhibited on the ends so it can be attached to the chamber. It burns on the outside, and burning area and thrust decrease as burning progresses.

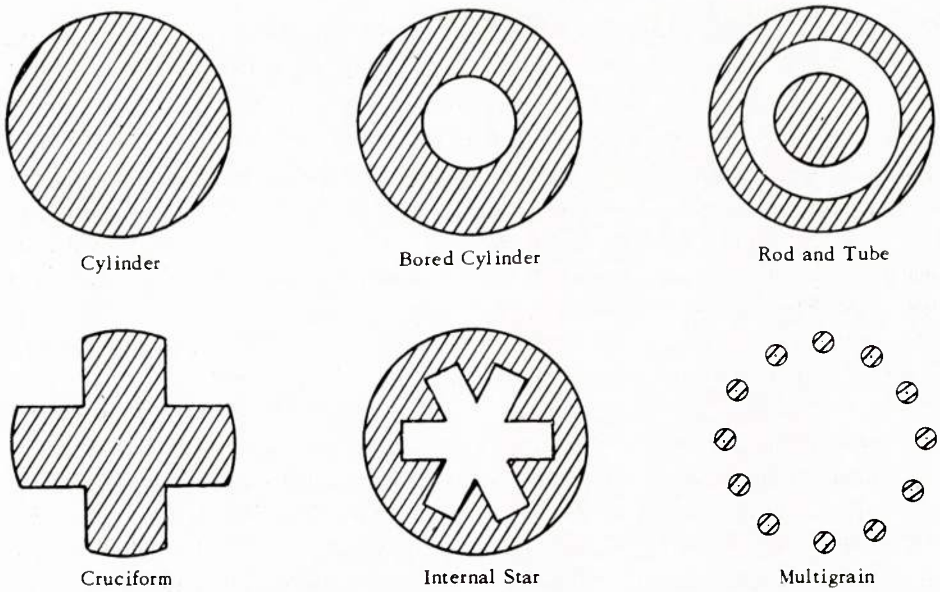


Figure 10. Typical solid propellant grain shapes shown in cross section.

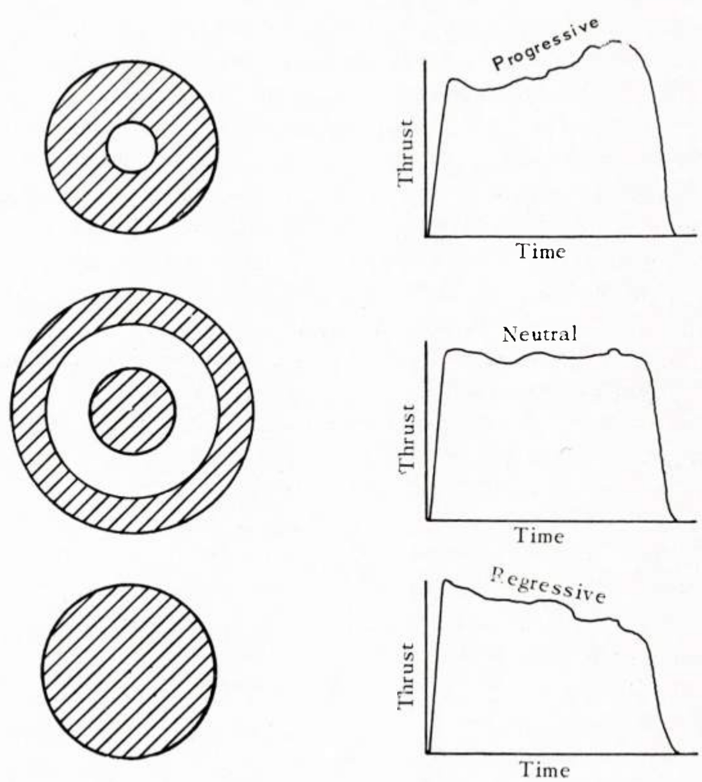


Figure 11. Progressive, neutral, and regressive burning grains.

Theoretical I_{sp} for typical solid propellants is shown in Table 3.

TABLE 3
Specific Impulse of Solid Propellant Combinations

Fuel base	*Oxidizer*	*Specific impulse (seconds)*
Asphalt	Perchlorate	200
Nitrocellulose and nitroglycerine	—	240
Polyurethane	Perchlorate	245
Boron	Perchlorate	270
Metallic hydride	Fluoride	300

Some of the earliest composite propellants had an asphalt and fuel oil base with about 75 percent potassium perchlorate oxidizer. Oil was added to the asphalt to keep it from becoming brittle and cracking at low temperature, but the resultant propellant became soft and began to flow at high temperatures. The specific impulse and operating range of these propellants was rather limited.

The double-base propellants use nitroglycerine and nitrocellulose with the proper additives. These propellants have a higher specific impulse than the asphalt-perchlorate propellants but are more sensitive to shock.

The polyurethane-ammonium perchlorate propellant is a typical example of the specific impulse of present-day synthetic rubber-based composite propellants.

In the past, the specific impulse of composite propellants has been improved by using a higher percentage of oxidizer. This was successful only as long as the propellant retained adequate structural properties, since the fuel-binder gives the propellant its mechanical strength. The specific impulse can also be improved if a light metal, such as aluminum, is added to the fuel. Light metals increase the combustion chamber temperature and lower the molecular weight of the combustion products. The result is a higher value of T_c/m and specific impulse. However, metal in the exhaust gases may cause erosion of the rocket nozzle.

Replacing the inert fuel binder with a high-energy fuel binder containing oxygen (double-base propellants) further increases specific impulse. Some of the propellants used today combine a double-base propellant with a composite propellant and a metal fuel. Propellants under development include boron and metallic hydride fuels in a suitable binder with a perchlorate or a fluoride oxidizer.

As a group, the solid propellants have a lower specific impulse than the liquid propellants, but they have the advantages of simplicity, instant readiness, low cost, potential for higher acceleration, and high density compared to the start-stop capability, high energy, regenerative cooling, and availability of the liquid propellants.

HYBRID PROPELLANTS—Some of the advantages of liquid and solid propellants can be combined in a hybrid rocket. In a hybrid engine, a liquid (usually oxidizer) is stored in one container while a solid (usually fuel) is stored in a second (See Figure 19). The separation of propellants in a hybrid eliminates the dependence of burning time on the grain temperature. The absence of oxidizer in the solid grain also improves its structural properties. The hybrid combines the start-stop advantages of liquid propellants with the high density, instant readiness, and potentially

high acceleration of the solid propellants. It is a relatively simple system with high performance. Solid fuel lithium, suspended in a plastic base, burned with a mixture of flourine and oxygen (FLOX) produces a theoretical I_{sp} of about 375 sec.

STORAGE IN SPACE—Propellant storage in space is one of the problems that must be considered in selecting the chemical propellants for space propulsion. The system may use the propellants intermittently over long periods of time, or it may store them for months prior to performing one maneuver. In space, the propellants no longer have the protection of the earth's atmosphere. They must perform in a hostile environment in which they are exposed to a hard vacuum, thermal radiation from the sun, energetic particles such as cosmic rays, and meteoroid bombardments.

Measures are necessary to protect propellants in order to prevent deterioration or loss by evaporation. Liquid propellants must be protected from evaporation and components must be leakproof. Solid propellants must be protected from radiation which affects their burning rate and physical properties. Radiation may also cause changes in liquid propellants. Propulsion systems must withstand temperature extremes when one side faces the sun and the other the coldness of space. They must also be protected from damage caused by micrometeoroids. These are a few of the problems caused by the space environment. Additional research and testing are needed to define all of the effects of space environment on propellants.

CHEMICAL ROCKET ENGINES

One basic method of improving I_{sp} by using more energetic propellants was discussed in detail in the previous section. Designing a better engine for a given propellant is the second basic method. The problem of achieving higher specific impulse generally is not solved by using only one method for improvement. Frequently both methods are used to produce the best results.

This section describes a few typical examples of liquid propellant, solid propellant, and hybrid propellant engines, and suggests a few of the many possible design improvements. The material is limited to the large and comparatively high-thrust rockets, although much of it applies to small rockets, such as those used for vernier and attitude control.

Chemical rocket engines may be classified by several different methods, one of which refers to the type of propellants used. The groups shown in Figure 12 should be remembered when comparisons are made and when methods for improving performance are discussed, because the different groups have different characteristics.

In general, when fast reaction is needed (as in military missiles) ease of handling and simplicity of operation are paramount. In such applications, solid propellant or storable liquid propellant engines are preferred.

For in-space maneuvering where restarts must be made, controllability coupled with high I_{sp}, is desired. The liquid propellant engines have these traits. Of course, compromises have been and will continue to be made based on attainable technology.

The hybrid engines are attempts to combine the advantages of both liquid and solid engines. Several small hybrid engines are in use.

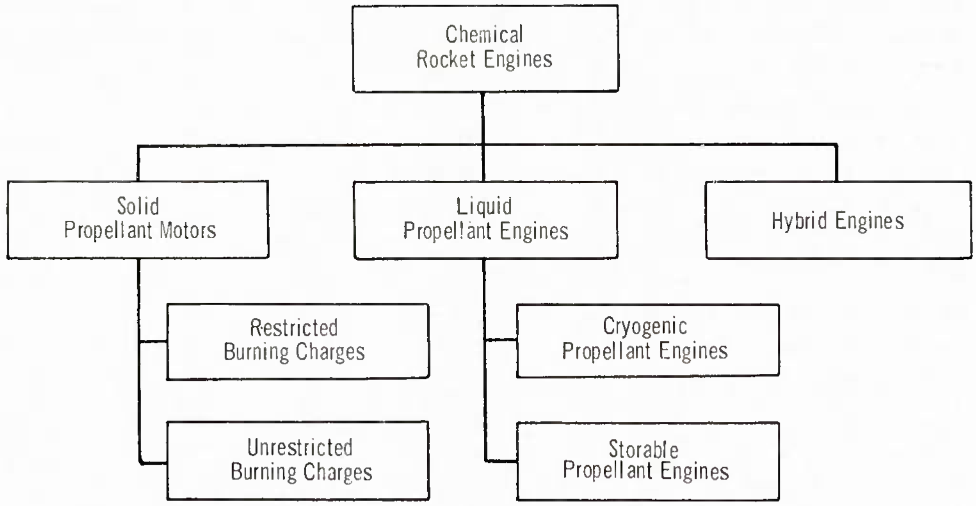


Figure 12. One method of classifying chemical propellant rockets.

Rocket engines are very lightweight for the amount of thrust they produce. Generally, a rocket engine will deliver about 100 pounds of thrust for every pound of engine weight, compared to the better jet engines which produce approximately 10 pounds of thrust per pound of engine weight. If the thrust of a vehicle's engines can be improved without increasing their weight, the propulsion system is improved.

In comparing two or more engines to determine which is best for an application, the entire vehicle and its mission must be evaluated. For example, consider two liquid propellant engines with *similar* thrust levels. The engine with higher I_{sp} would be more efficient and have higher vehicle velocity potential. When engines with similar operating characteristics, propellants, and thrust levels are compared, higher I_{sp} results from better propellant utilization.

A comparison might also be made between two similar liquid engines with similar propellant consumption rates, but different thrust levels. The engine with higher thrust yields higher I_{sp} and is more efficient. Designers try to achieve the highest possible I_{sp} for any given rocket because more I_{sp} yields increased velocity and payload.

Liquid Propellant Engines

A liquid propulsion system consists of propellant tanks, propellant feed system, thrust chamber, and controls such as regulators, valves, and sequencing and sensing equipment. The propellants can be monopropellants, bipropellants, or tripropellants, and may be either storable or cryogenic fluids.

The least complex of these is one designed for monopropellants. Here there is only one propellant tank, a single feed system (usually pressure-fed), and a comparatively simple injector (since mixing of fuel and oxidizer is not required). Monopropellant rockets are in limited use today but do not yet develop high

thrust. However, the simplicity of monopropellant engines makes them adaptable and frequently desirable for limited use for attitude control or for small velocity corrections in deep space.

The liquid bipropellant system in common use is more complex. The basic components are shown schematically in Figure 13. Note that two tanks, two feed systems, and multiple injectors are required.

Some bipropellant systems use pressure-fed propellant flow. Here propellants are forced from tanks to engine by pressurizing the tanks with an inactive gas, such as nitrogen or helium. A pressurizing gas can also be created by igniting either a solid propellant grain or some of the vehicle propellants in a gas generator designed for this purpose.

Bipropellants may also be fed to the engine by pumps and gravity, as was done in some earlier booster engines. However, a combination of pumps and pressurizing tanks to provide positive pressure to the pumps feeding the engines is the most commonly used method today. The bipropellant system is complex

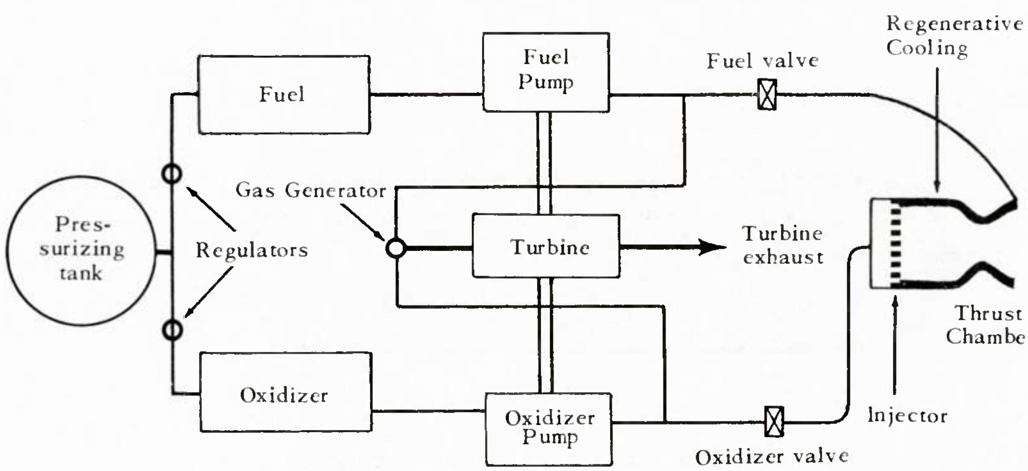


Figure 13. Schematic reproduction of a liquid bi-propellant system.

because of the multiple pumps, the need to maintain the correct oxidizer-fuel mixture, the effect of the injector design upon stable combustion, and the need for thrust chamber cooling.

A centrifugal pump driven by a gas turbine is commonly used to pump the propellants through the injector into the combustion chamber. Gases to drive the turbine are supplied by a separate gas generator, or they are bled from the combustion chamber. The development of reliable turbopumps has presented severe challenges in design, materials, testing, and operational use. In some instances, more than 50% of the design effort for an engine was devoted to the turbopump.

Turbopumps must also pump fuel and oxidizer simultaneously at different rates and be able to withstand high thermal stresses induced by the 1500° F. turbine gases while pumping cryogenic fluids with temperatures as low as −423° F. (liquid hydrogen). Adequate seals must be continuously maintained even at

these temperature extremes, since the propellants would explode if they were to come in contact inside the pump. Pump design is also critical because the pumps must develop high propellant pressure. Higher combustion chamber pressure means higher I_{sp}, and pump outlet pressure must be higher than chamber pressure if propellants are to flow into the chamber.

Since the components and controls of the liquid engine can be designed for individual control, the potentials of throttling and multiple restart make these engines attractive for in-space maneuvering.

Solid Propellant Rocket Motors

Solid propellant rocket motors have been in use for thousands of years. History tells of the ancient Chinese using skyrockets for celebrations as well as for weapons. The "rockets' red glare," as used in the National Anthem, indicates the use of rockets during the War of 1812. JATO units, to decrease aircraft take off roll or as take off assist units for lifting heavy loads, are familiar to most Air Force personnel.

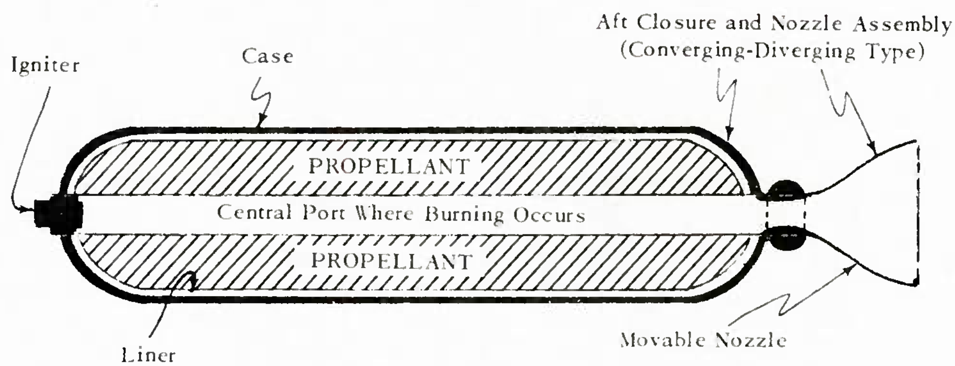


Figure 14. A solid propellant rocket motor.

The solid propellant rocket is comparatively simple. The major components are: the case, which holds the propellants and is the combustion chamber; the igniter; and the converging-diverging nozzle (Fig. 14). Because of its simplicity, the solid motor is inherently more reliable and cheaper to produce than the liquid rocket engine. However, these solids have presented problems, and have a lower I_{sp} than the liquid engines.

As mentioned earlier in the propellant section, the use of additives, new chemicals, and design of high volumetric loading propellant grains is improving the I_{sp} of solids. The other approach for increasing performance is to increase the mass ratio of the motor. Much effort has been expended in this area by designing cases that are lightweight and stronger. Research seems to point to two possible solutions: either make lightweight but strong cases of metals such as titanium, or design filament-wound cases of fiberglass or nylon tape impregnated with epoxy-type glues.

The filament-wound cases can be made even lighter if reinforced propellant grains are designed to assist in supporting the vehicle. Reinforced grains are formed by molding the propellant around aluminum or other metal additive wires. These reinforcing materials are consumed during combustion. Reinforced grain motors are usually regressive burning so that combustion chamber pressure will decrease near the end of burning to allow the use of very lightweight cases.

Because of the relative simplicity of solid motors, they can be made in a variety of sizes from very small attitude control and docking motors with 0.1 lb of thrust to 260 in. diameter motors which produce up to 6 million pounds of thrust.

Thrust Vector Control

Attitude and directional control of a rocket is accomplished by moving the engines or deflecting the exhaust gases. This is called thrust vector control (TVC). The thrust vector for the main and vernier engines is controlled by using flexible mountings or gimbals for the engines. Hydraulic or pneumatic cylinders deflect the engines to change the flight path.

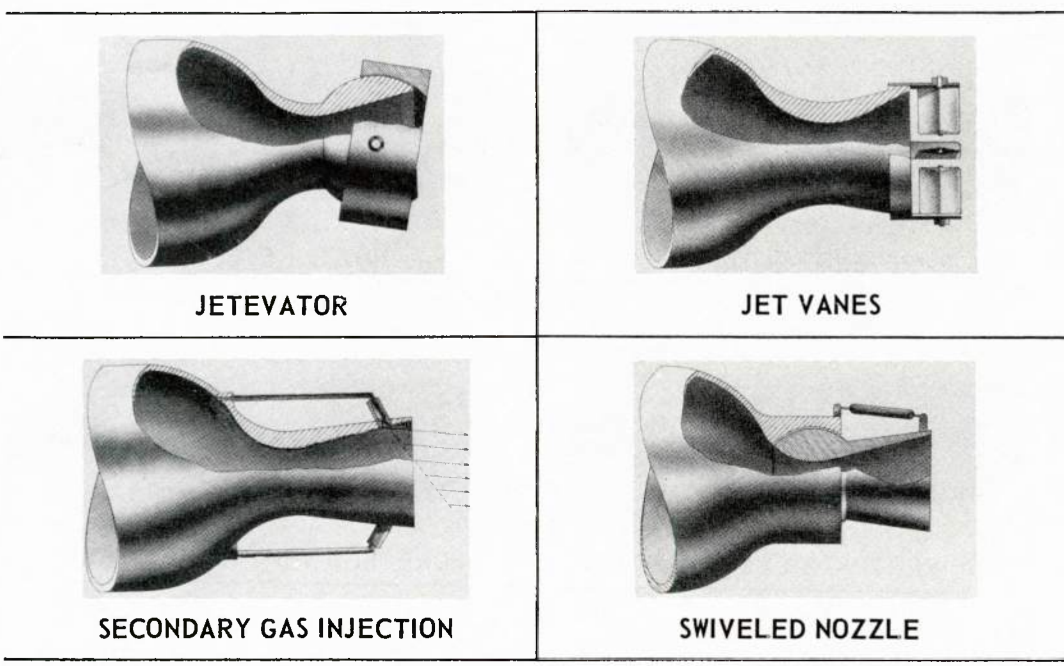


Figure 15. Methods of thrust vector control (TVC).

In solid motors it is impractical to deflect, or swivel, the entire motor because this amounts to moving the entire stage. The thrust vector in solid motors must be controlled at the nozzles. A few of the more common methods employed are: *movable nozzles; controllable vanes* in the nozzles; *jetavators* (slip-ring or collar at the nozzle exit); or *injection of fluid* (hot or cold gases) into nozzles to deflect the exhaust flow and accomplish flight path or attitude changes (see Fig. 15).

Thrust Termination

Thrust termination in liquid rockets is comparatively simple because it is only necessary to stop the flow of propellants. Since this can be done in a reproducible sequence, the residual thrust generated during, and shortly after, engine cut-off will be a known quantity. Liquid engines can be designed so that the thrust level can be changed within limits by varying the rate at which propellants flow into the engine.

In solid motors thrust termination is also simple (Fig. 16). If the combustion chamber pressure is reduced below the critical pressure value for that particular motor, it will, in effect, blow itself out. Blowing of nozzles, blowing out the aft end of the motor, or using thrust termination ports to vent the pressure out the sides of the case are some of the common methods used to reduce the chamber pressure and terminate thrust. However, the thrust termination is not instantaneous. Low levels of thrust may continue for several minutes because of residual burning in the remaining solid grain. The residual burning characteristics of each type of motor must be considered when a very accurate cut-off velocity is needed.

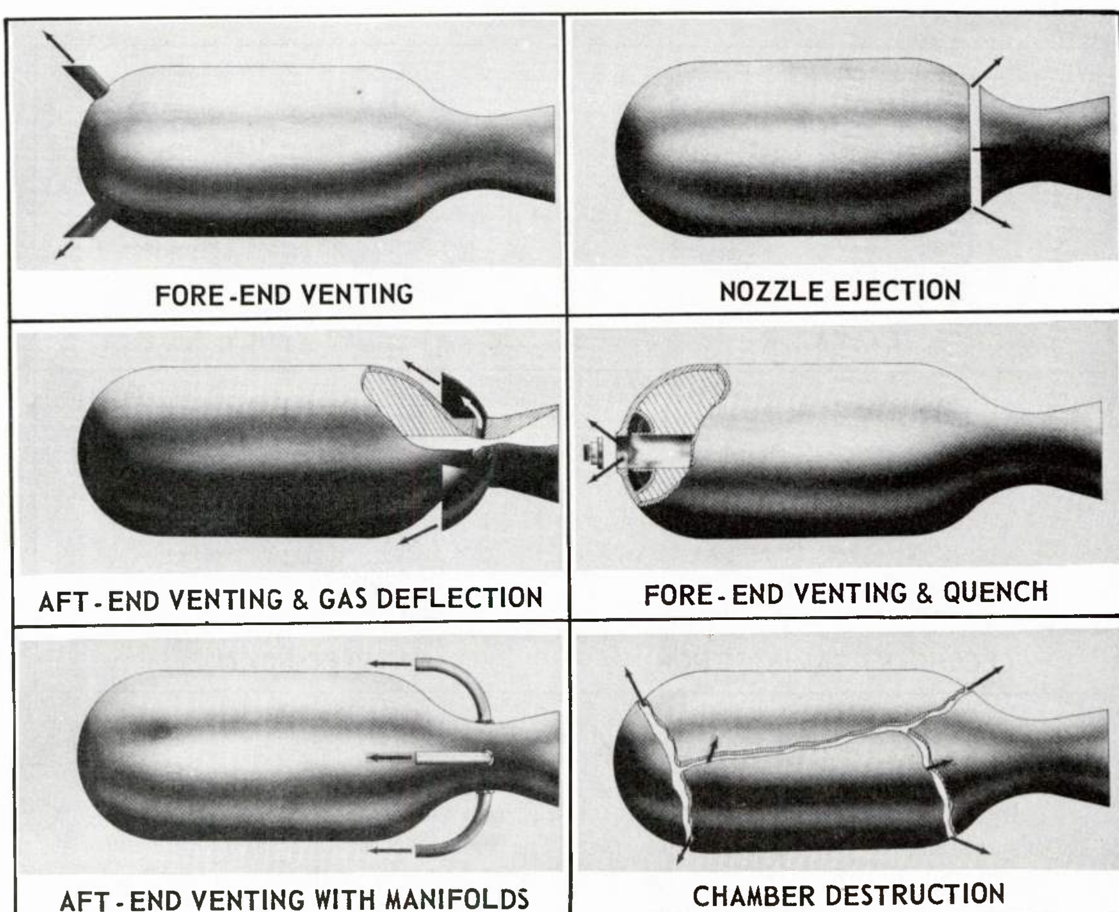


Figure 16. Methods of thrust termination.

Engine Cooling

The combustion chamber of a rocket engine contains gases at temperatures in excess of 5000° F. Large amounts of heat are absorbed by the walls of the combustion chamber and nozzle. Some provision must be made to dissipate heat; otherwise, as the heat is absorbed, the engine wall temperature increases, and the strength of the structural materials decreases. Since the amount of heat absorbed is highest in the nozzle throat, special attention is given to this region.

Most liquid engines are either partially or completely cooled because uncooled liquid rockets are not fired for periods greater than 25 sec. If combustion temperatures were low, it might be possible to use an uncooled engine for longer periods of time; but such is usually not the case in large engines. Cooling methods include regenerative, water, film, sweat (transpiration), radiation, and ablative cooling.

Regenerative cooling circulates fuel or oxidizer through small passageways between the inner and outer walls of the combustion chamber, throat, and nozzle. The heat removed cools the engine and increases the energy of the propellant before it is injected into the combustion chamber. The energy added to the propellant slightly increases the velocity of the exhaust gases and improves engine performance.

Water cooling is regenerative cooling, except that water circulates instead of the fuel or oxidizer. Water cooling is widely used in rocket engine static test stand firings, but not on the flight vehicle.

Film cooling provides a thin fluid film to cover and protect the inner wall of the engine. A protective film is formed on the inner wall by injecting small quantities of the fuel, oxidizer, or a nonreactive fluid at a number of points along the hot surface. The fluid flows along the wall and absorbs heat by evaporation. Film cooling can be used with regenerative cooling for critical parts of the engine where regenerative cooling alone is not sufficient.

Sweat or transpiration cooling uses a porous material for the inner wall of the engine. Coolant passes through this porous wall and is distributed over the hot surface. It is difficult to distribute the coolant uniformly over the surface because the combustion gas pressure decreases between the combustion chamber and the nozzle exit. Manufacturing porous materials so that they are uniform throughout the engine is difficult, and this difficulty limits the use of this method of cooling.

Radiation cooling removes heat from the engine and radiates it to space. Some current liquid engines use a regeneratively cooled nozzle to the 10:1 expansion ratio point and a radiation-cooled extension from that point to the end of the nozzle. This type is lighter than it would be if regenerative cooling were used for the entire nozzle. Current research includes investigating radiation-cooled combustion chambers, nozzles, and nozzle extensions.

Ablative cooling involves coating the surface to be cooled with a layer of plastics and resins. The coating absorbs heat, chars, and then flakes off, carrying heat away from the subsurface being protected.

The cooling methods most commonly used with liquid engines are regenerative cooling, film cooling, or a combination of the two. Since cooling a solid motor would require additional cooling equipment and materials, other protective means have been devised. Inserts of high temperature metal alloys, graphite or ceramics have been used with success. Using multiple nozzles and ablative materials in the nozzle also keeps temperatures within working limits. Heating is also controlled through grain design and use of liners or inhibitors in the case. Combustion gas temperatures in solid motors are generally lower than those in liquid engines.

Nozzles

The development of lighter weight, shorter nozzles has been in progress for several years. There are several reasons for this: (1) to reduce dead weight; (2) to reduce the interstage structure of multistage vehicles; and (3) to optimize the expansion ratio at all altitudes. Several improved nozzles are shown in Figure 17.

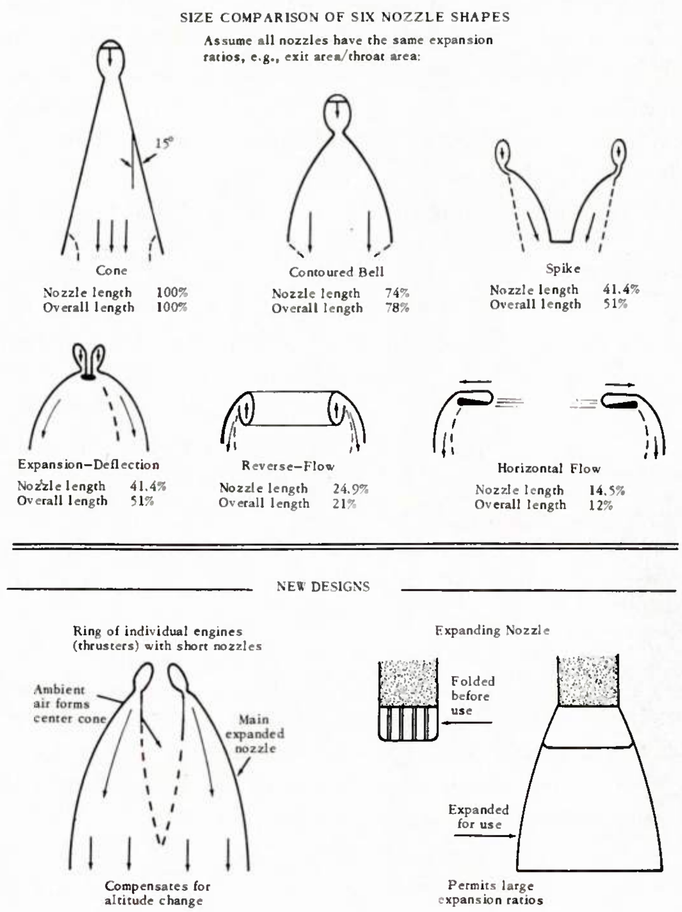


Figure 17. Nozzle design schematics.

Improvements

Several recent concepts have been used to develop bigger and better chemical rockets. Two such developments are the segmented solid and the hybrid rocket.

Solid motors can be segmented or monolithic. The segmented motor is made by stacking separate grains (or segments) together to give the desired thrust level. Using segments greatly eases casting, inspection, and transportation problems. Special methods of assembly insure that cracks do not form between segments and cause the motor to explode.

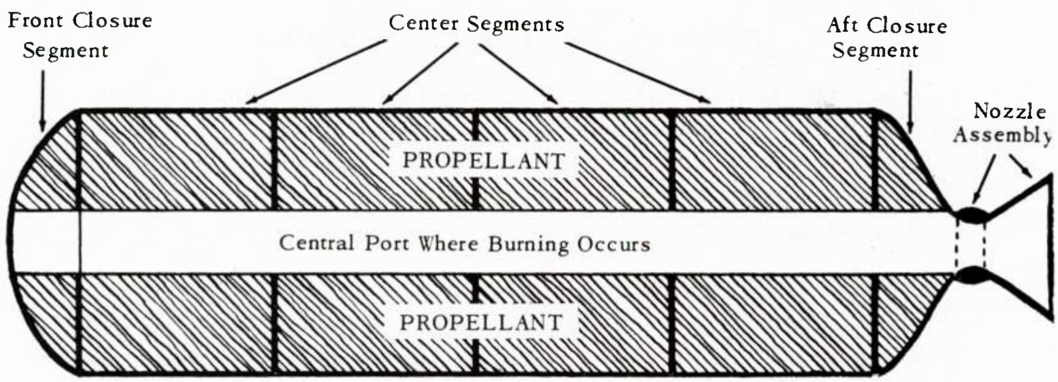


Figure 18. A segmented solid propellant rocket motor.

The segmented solid idea (Fig. 18) has been used to build large diameter motors. These are manufactured in four basic units or "building blocks": the front closure segment, the center segments, the aft closure segment, and the nozzle assembly. The average thrust level of the rocket can be changed by varying the number of center segments, thus controlling the burning area. For example, five-segment, 120 in. diameter, solid motors are used as boosters for the Titan III C with each motor developing 1.2 million lbs of thrust. A seven-segment motor of the same diameter develops 1.6 million lbs of thrust. Above 156 in. diameter, the motors are so large that they will probably be manufactured or cast at the launch site, or transported by water. A monolithic (one piece) motor would be the simplest construction in this case.

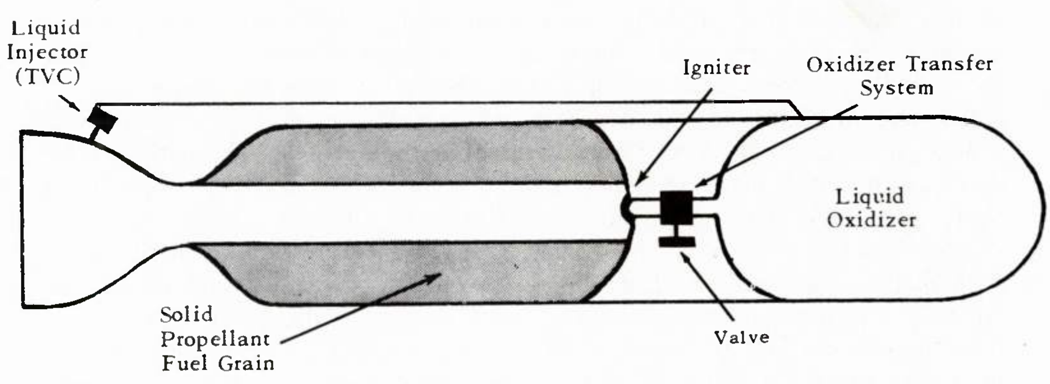


Figure 19. Hybrid rocket engine.

In the hybrid engine (Fig. 19), liquid and solid propellants are used in one engine. The hybrid represents a compromise. It attempts to take advantage of the simplicity and reliability of solids; the higher performance and throttleability of liquids; storability and quick reaction time of solids; and, it reduces the hazard of having both fuel and oxidizer mixed together in one case.

In summary, present trends in liquid rocket engines include very high chamber pressures, high-energy propellants, and simplification of hardware. Solid rockets are being improved by using lighter weight cases, thrust vector control, and uncooled nozzles.

ADVANCED PROPULSION TECHNIQUES

As mission requirements approach or exceed the limits of chemical rockets, new propulsion techniques must be investigated. The new techniques will obey Newton's laws just as contemporary techniques do, but they will use different energy sources and hardware to produce the propulsive force. This force will still be the reaction of the vehicle to mass being ejected at high velocity.

Need for Advanced Designs

The need for advanced designs becomes readily apparent as velocity limitations of chemical rockets are considered. These limitations are shown in Figure 20 for single and multistage vehicles. The graphs include both LOX/RP–1, as well as the higher energy combination LOX/LH_2. *Payload fraction* is the payload weight divided by the gross vehicle weight.

Since there is a practical limit to the number of stages that can be used to increase mass ratio, consider that I_{sp} is proportional to T_c/m. Remember also that T_c and m interact and influence each other in a combustion process. If T_c and m were independent, they could be varied so that the resulting ratio would be higher. The value for m is about 20 in present chemical engines. If m is reduced significantly, a higher performance engine results.

Nuclear Rocket

One program that has followed this approach is the nuclear rocket. It has increased the theoretical I_{sp} to more than 800 sec—double that of current chemical engines. A schematic reproduction of a typical nuclear, solid core, thermal reactor engine is shown in Figure 21. Since liquid hydrogen is used as the propellant, thrust levels are comparable to upper stage chemical engines using hydrogen.

The nuclear rocket was initiated as a joint *A*tomic *E*nergy *C*ommission-USAF project but became an AEC-NASA program of two phases. The early *reactor core design* phase used the Kiwi A and B reactors and was completed in September, 1964. The current phase involves developing the *flight configuration* engine known as NERVA, "*N*uclear *E*ngine for *R*ocket *V*ehicle *A*pplication." Preliminary flight tests are expected to be ballistic trajectory flights with the NERVA functioning as an upper stage main engine. Many areas of study and research are necessary before the nuclear rocket is ready for production. Among these are starting at altitude, and radiation and neutron heating which may

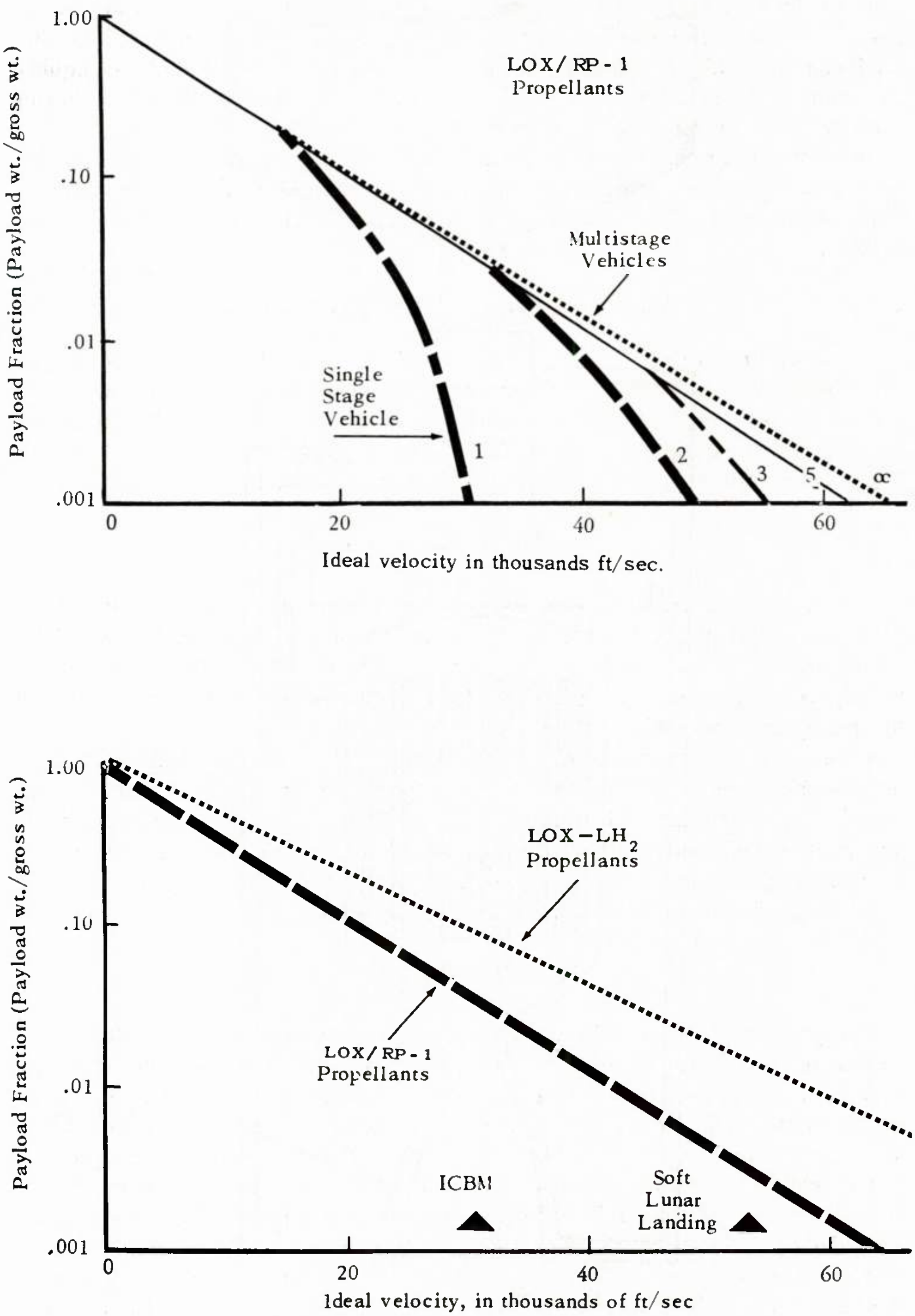


Figure 20. Payload fraction versus ideal velocity.

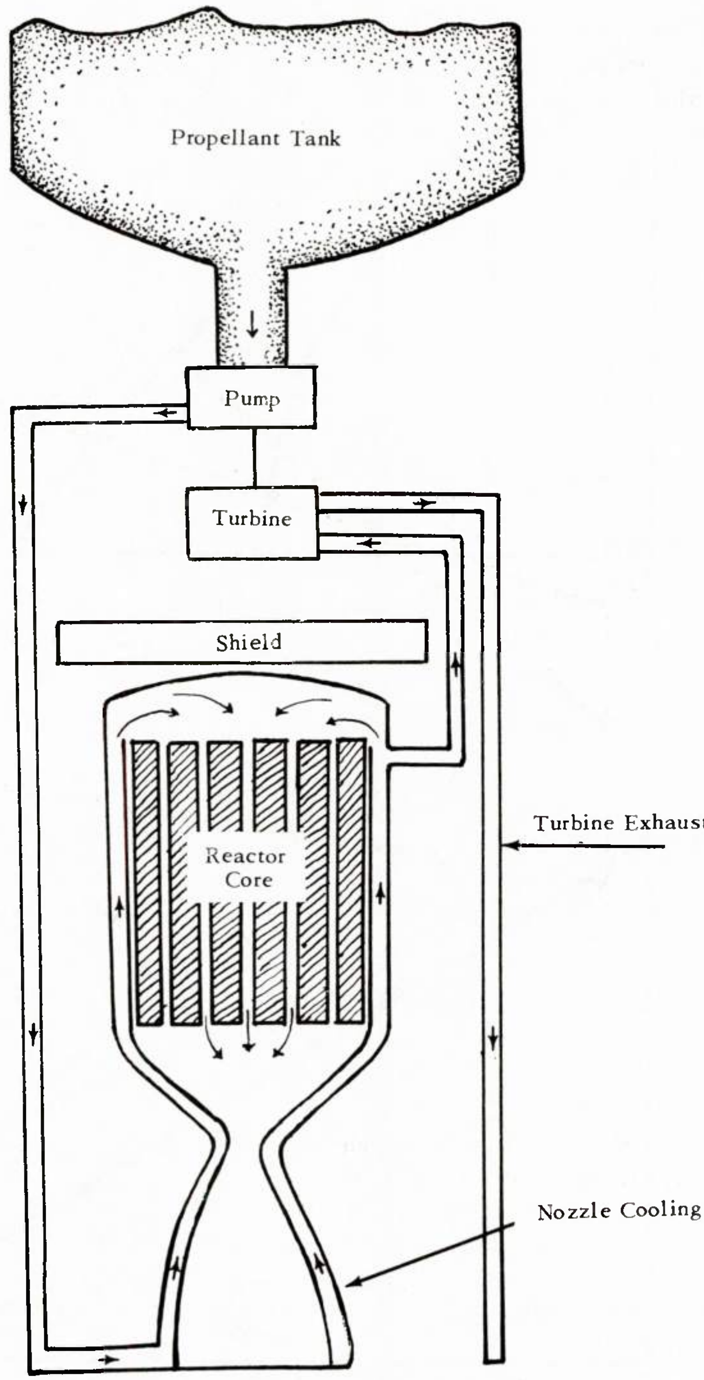


Figure 21. A nuclear rocket.

affect the payload, propellants, and structural materials. A more advanced increased density reactor called PHOEBUS has been tested, and a flyable engine is now being designed.

One theoretical improvement being considered is a high density reactor using fast neutrons. This type of reactor is expected to produce higher performance levels in a

smaller package than the thermal (or slow) reactors mentioned above. Another improvement that may prove feasible at some time in the future is a gas core reactor, in which the operating temperature (T_c) could be much higher. This increase in temperature would occur because of the elimination of the solid core or fuel elements used in slow and fast reactors. These structural elements are temperature limited.

The theoretical, comparative performance of thermal reactor engines, fast reactor engines, and chemical engines is shown in figure 22.

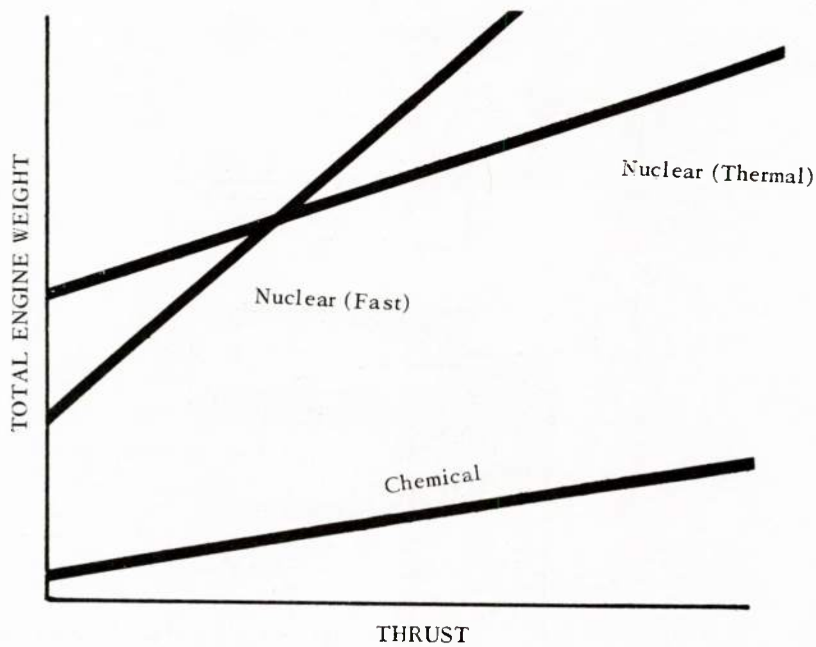


Figure 22. Thrust vs. weight of a rocket.

Low Thrust Rockets

Chemical engines produce high thrust for a short time (minutes), and nuclear reactor engines yield high thrust for hours. Each produces acceleration that stops when the propellant is exhausted.

There are other engines that produce only small amounts of thrust, but they do so for months, or even years. An unbalanced thrust acting for long periods can produce final velocities much higher than those produced by chemical or reactor engines.

The *radioisotope heat cycle* and the *electric* engines produce thrusts measured in micro and millipounds, specific impulses from 700 to 30,000 seconds, and operating times ranging from days to years. These engines cannot lift themselves from the earth, but they can move large payloads through space.

The *radioisotope heat cycle* engines use high energy particle sources such as plutonium and polonium. The particles are stopped by the walls of the isotope container thereby converting their kinetic energy to heat. This heat is used to

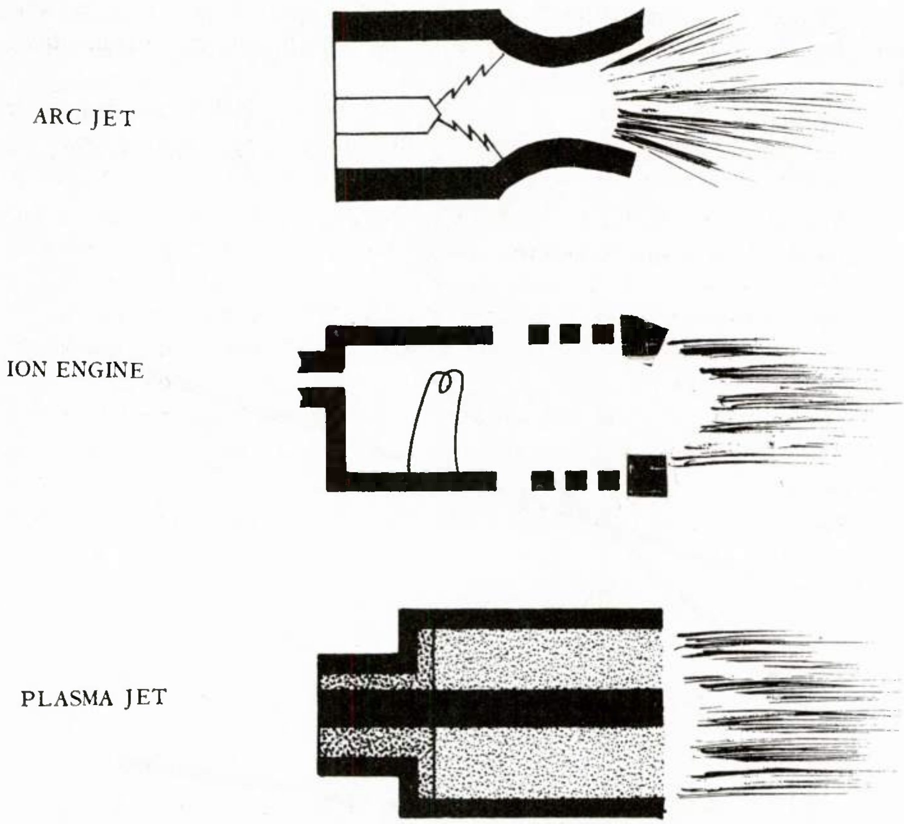


Figure 23. Electric rocket thrust chambers.

raise the temperature and pressure of a propellant which is then expelled through a nozzle.

The DART (*D*ecomposed *A*mmonia *R*adioisotope *T*hruster) is such a device. Here, ammonia is heated by 238 plutonium dioxide to produce 0.1 lb of thrust, and an operating time of at least one year is being sought.

There are basically three types of *electric* engines: arc jet, ion engine, and plasma jet. These are shown schematically in Figure 23.

The *arc jet* uses an electric arc to heat the propellant which is accelerated thermally and ejected as a high velocity plasma from a conventional nozzle.

The *ion engine,* is an electrostatic device which removes electrons from the propellant atoms to form positive ions. These ions are electrostatically accelerated and ejected to produce thrust. Electrons must then be added to make the exhaust electrically neutral to prevent accumulation of a negative charge on the vehicle.

The *plasma jet,* uses electromagnetic force to accelerate and eject the propellant in a plasma form to provide thrust. Electromagnetic force is necessary, since a plasma is a substance which is ionized but is electrically neutral (a plasma consists of an equal number of positive and negative ions, and therefore it has no net charge).

This is, of course, a very simplified presentation. The theory and associated equations for these devices, particularly those for the plasma jet, become quite complex.

The three basic types of electric engines have many subdivisions based on such considerations as design, the method of transferring energy to the propellant, thrust, propellant consumption rates, and I_{sp}. Electric rockets are expected to yield specific impulses of 2,000 to 30,000 sec, or more.

When electric engines are considered, several significant factors must be recognized. All such devices require considerable electrical power, perhaps for years, if they are used for primary propulsion. Only nuclear-electric generators can supply enough power for these long periods of operation. Figure 24 shows a nuclear-electric power and propulsion system. Electric engines have low thrust levels on the order of 0.1 to 0.001 lb. Design improvement will increase these levels. Since electric engines have low thrust-to-weight ratios, they must be used in space where there are weak gravitational fields.

These engines have great potential for some missions. A typical ion rocket uses

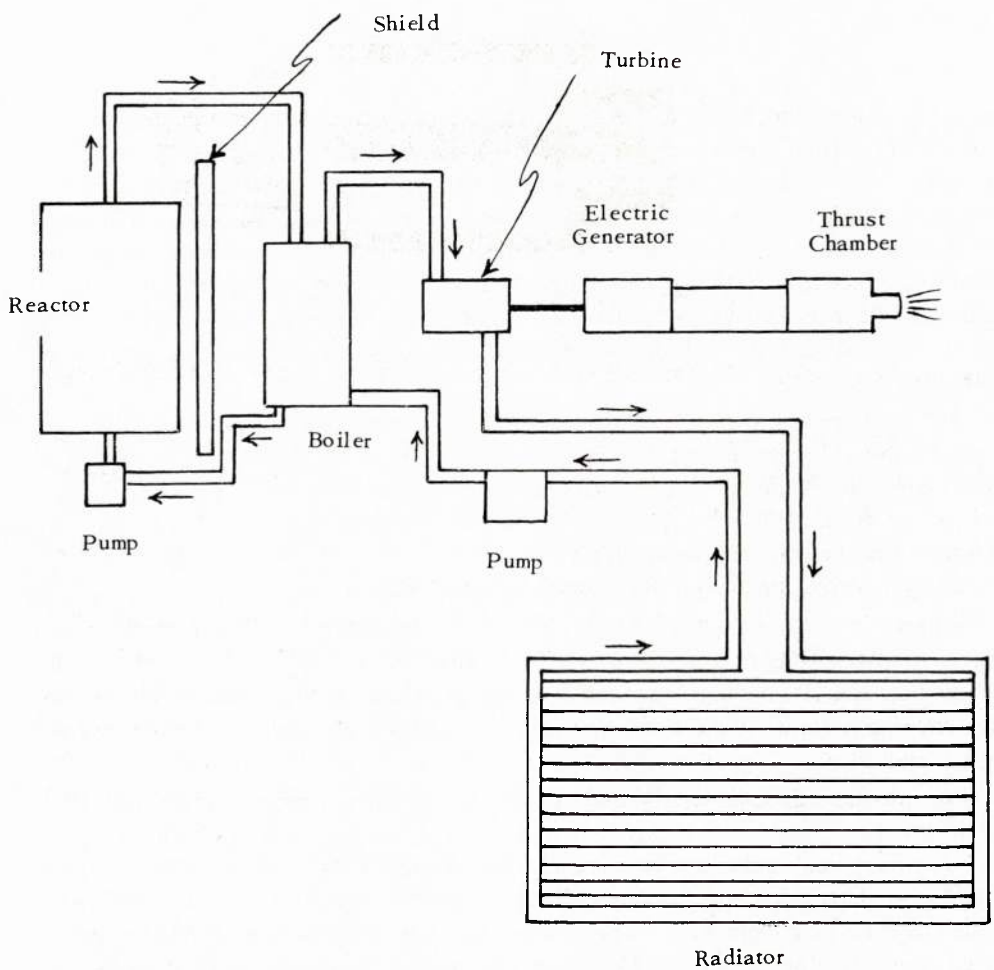


Figure 24. Nuclear-electric power and propulsion system.

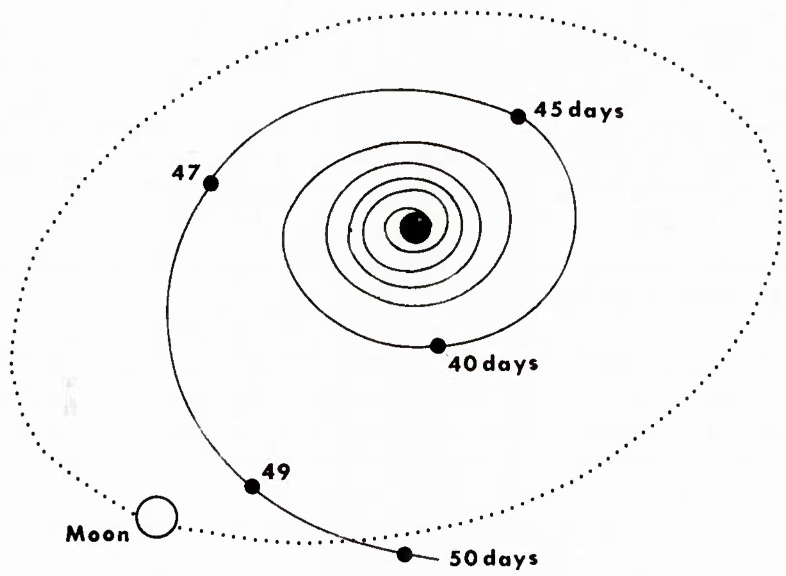


Figure 25. Escape trajectory.

some working fluid (such as cesium) which is vaporized and routed to the "engine." The fluid has electrons stripped from its atoms by passing vaporized cesium over a heated platinum grid. This stripping action leaves positively charged atoms or ions which are accelerated by an electrostatic field and ejected to produce thrust. The stripped-off electrons (negative ions) are also accelerated and introduced into the positive ion stream producing an electrically neutral exhaust. If a difference in potential of about 30,000 volts exists between the platinum grid and the cathode acccelerator, ion velocities in the exhaust exceed 500,000 ft/sec.

An electrically powered vehicle with a 2/1 mass ratio and a 20,000 sec specific impulse can be useful in space. Starting from a 200 NM earth orbit, this vehicle might require 2½ to 3 years to reach its peak velocity, but that velocity would be 80 to 100 miles per sec. Compare this with the 5 to 7 mile per sec of our current earth-orbiting satellites.

From the standpoint of a practical mission, the electrically powered vehicle can move a sizable payload to the vicinity of the moon in approximately 50 days, following the trajectory shown in Figure 25. If long flight times can be tolerated, electric engines can move large payloads out to Mars and Venus. If long flight times are not feasible, either chemical or nuclcar reactor engines must be used.

For missions beyond Mars and Venus the electric engines come into their own and may easily excel chemical and nuclear reactor rockets, both of which achieve their peak velocities in a rather short powered flight phase. On the other hand, the electric systems continue to accelerate for days, even weeks, and attain much higher final velocities. This means there is a mission crossover point in deep space (in the vicinity of Mars and beyond) at which the electric propulsion

TABLE 4
Mission Summary

TYPE	*MISSIONS*
CHEMICAL PROPULSION	Manned missions near Earth and Moon return. Instrumented probes to Venus and Mars.
NUCLEAR PROPULSION	Heavy payload manned missions in 1970's to Moon, Venus, and Mars and return.
ELECTRIC PROPULSION	Potential for very heavy payloads from Earth orbit to vicinity of Mercury, Jupiter and Saturn (low gravitational field applications only).

systems have the advantages of both shorter flight times and larger payloads. This is shown in Table 4.

Future Propulsion Concepts

Discussion in this chapter has thus far included propulsion systems in production (chemical) and systems in active research and development programs (nuclear and electric). Now concepts for possible future propulsion will be considered.

ADVANCED NUCLEAR—An advanced nuclear concept is one in which propulsive power is derived from a series of controlled nuclear explosions. The force of the explosions impinges on the pusher plate which, in turn, transfers the energy to the main vehicle through a shock absorber system. The shock absorber smooths out the impulses and limits the "G" loading on the main vehicle to an acceptable level. This concept, shown in Figure 26, would provide a controllable specific impulse, and could propel a large payload out to Mars, Venus, or beyond. Payload fractions as high as 45% are possible with this system.

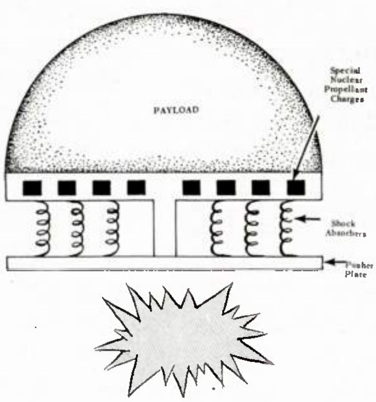


Figure 26. Advanced nuclear impulse concept.

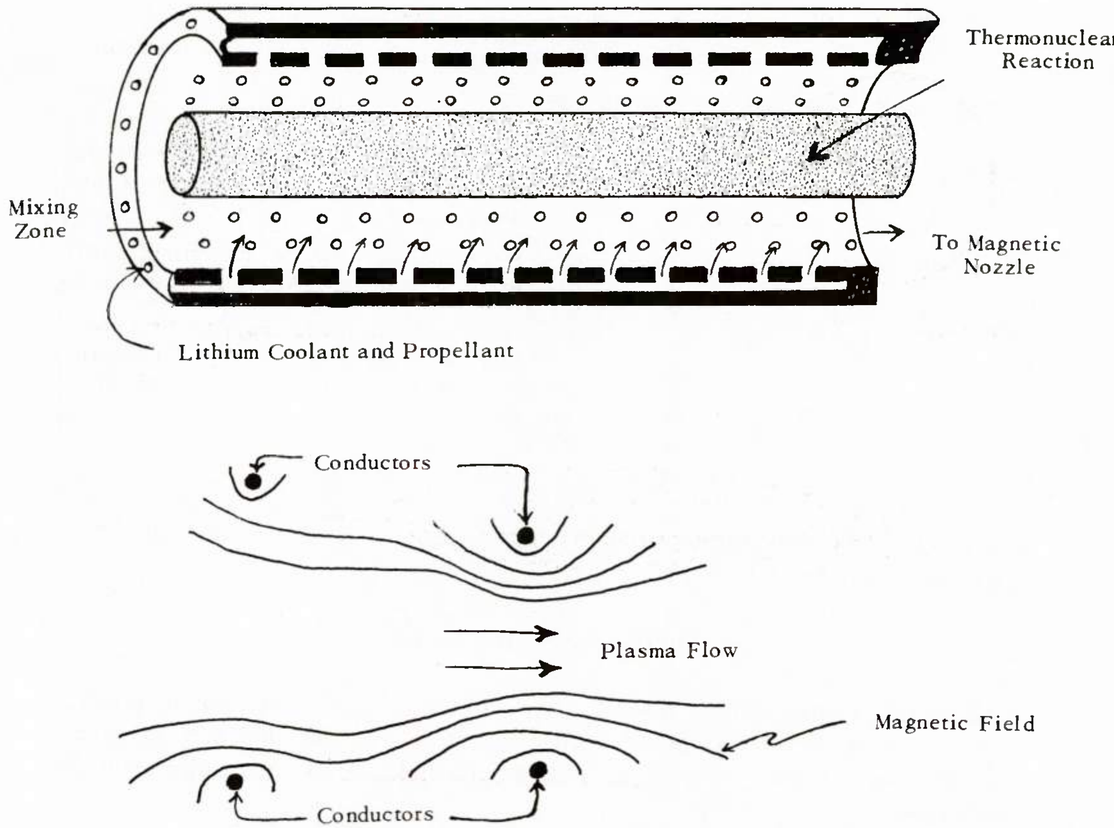


Figure 27. Fusion propulsion concept.

FUSION PROPULSION—Another advanced concept would use a fusion reaction contained by an electromagnetic field (Fig. 27).
The field also heats and controls the reaction, and controls the ejection of mass from the exhaust nozzle. The propellants are two isotopes of hydrogen. Deuterium (H_2) and tritium (H_3) are proposed because they can sustain a fusion reaction when sufficiently heated. This concept may yield a theoretical specific impulse of one million sec.

Scientists have proposed using a porous envelope through which lithium coolant (and propellant) would be forced into the reaction. The lithium would enter into the reaction and be expelled with the other exhaust products. When this is done, more mass flows, lowering the specific impulse to a value of about 4,000 sec, *but* the thrust level rises to a very high value.

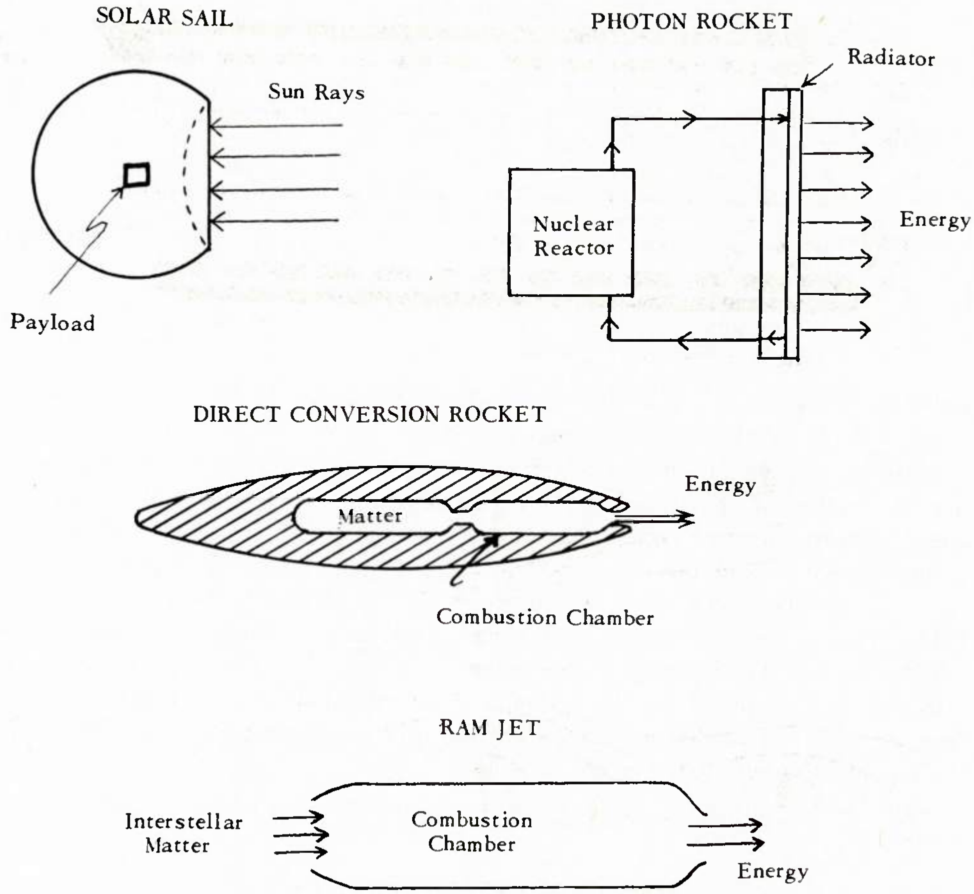


Figure 28. Photon propulsion concepts.

PHOTON PROPULSION—Another idea that may prove feasible in the future is the photon concept. Four approaches for using photon propulsion are shown in Figure 28.

SUMMARY

This chapter presented the basic laws of rocket operation, and some of the terminology and definitions pertaining to rocket engine performance. It also discussed the interaction of design and sizing parameters of rocket vehicles, and methods of increasing rocket vehicle performance. Some of the advanced pro-

TABLE 5

Propulsion Systems

Type	*Specific Impulse, sec.*
High Thrust — To — Weight Ratio	
Chemical	To 600
Nuclear	800–2,000
Low Thrust — To — Weight Ratio	
Electric	2,000–30,000
Fusion	1,000,000
Photon	30,000,000

pulsion concepts were also reviewed. A comparison of the relative specific impulse values for all of these is shown in Table 5.

Although the various types in Table 5 are compared on the basis of specific impulse, it alone does not present the complete picture. Only chemical propellants and high thrust nuclear rockets have thrust-to-weight ratios that are sufficient to permit launches from the earth. Of course, this also means a launch potential from other bodies, such as the moon, Mars, and Venus which have gravitational fields. Others (electric, and photon) are applicable to missions in low gravitational fields. These fields could be in deep space or between one planet's orbit and another planet's orbit. Again, a vehicle with chemical or nuclear reactor engines would be needed to travel from the orbital vehicle to a planet or moon surface, and to return to the orbital vehicle.

Various applications of engines for the immediate future are summarized in Figure 29.

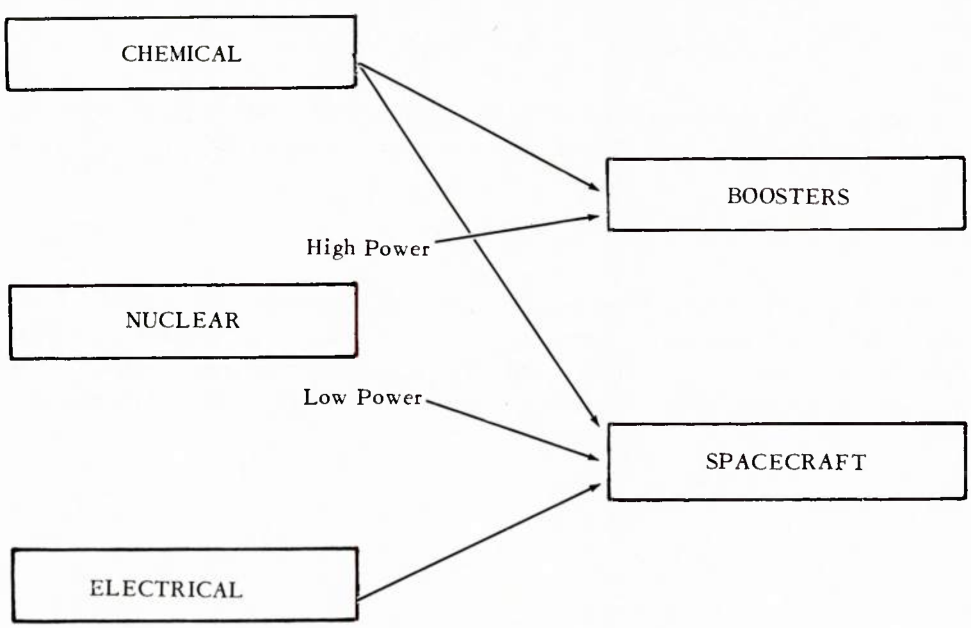


Figure 29. Types of propulsion and their applications.

BIBLIOGRAPHY

BERNATOWICZ, D. T., GUENTERT, D. C., and KLANN, J. L. "Space Powerplant Needs and Selection," *Astronautics and Aerospace Engineering,* I, 4 (May 1963), 22–26.

FAIRES, V. M. *Thermodynamics.* New York: Macmillan, 1962.

FRIEND, E. and HOYER, S. "Ion Engine Spacecraft Design Study and Mission Analysis," Paper No. 67-87 in AIAA Aerospace Sciences Meeting. 5th, New York, 1967. *AIAA Papers.* New York: American Institute of Aeronautics and Astronautics, 1967.

HESSE, W. J. and MUMFORD, N. V. S., Jr. *Jet Propulsion for Aerospace Applications.* New York: Pitman, 1964.

PEDERSEN, E. S. *Nuclear Energy in Space.* Englewood Cliffs, N. J.: Prentice-Hall, 1964.

SCHWENK, F. C. "Propulsion for Space: The Nuclear Rocket," *TRW Space Log,* VII, 1 (Spring 1967), 2–24.

SHAPIRO, A. H. *The Dynamics and Thermodynamics of Compressible Fluid Flow.* New York: The Ronald Press Company, 1953.

STAMBLER, I. "The Future for Large Solids," *Space/Aeronautics,* XLIV, (7 December, 1965), 74–81.

STUHLINGER, E. *Ion Propulsion for Space Flight.* New York: McGraw-Hill, 1964.

SUTTON, G. P. *Rocket Propulsion Elements.* New York: Wiley and Sons, 1964.

ZUCROW, M. J. *Aircraft and Missile Propulsion.* New York: Wiley and Sons, 1958.

PROPULSION SYMBOLS

a—acceleration
A_e—nozzle exit area
A_t—nozzle throat area
C_p—constant pressure specific heat
C_v—constant volume specific heat
F—thrust
g—acceleration of gravity
I_d—density impulse
I_{sp}—specific impulse
I_t—total impulse
k—ratio of specific heats
1n—natural logarithm
m—molecular weight
M—mass
$\dot{M}$—mass rate of flow
MR—mass ratio
p—electric power
P_c—combustion chamber pressure
P_e—exhaust pressure
P_o—ambient pressure
Q—reactor thermal power
r—mixture ratio
t_b—burning time
T_c—combustion chamber temperature
T_e—exit temperature
T_i—inlet temperature
V—volume
Δv_a—actual velocity change
Δv_i—ideal velocity change
Δv_l—velocity change losses
v_e—nozzle exit velocity
v_r—earth surface velocity
W—weight
$\dot{W}$—weight rate of flow
$\dot{W}_f$—weight rate of fuel flow
$\dot{W}_o$—weight rate of oxidizer flow
W_p—propellant weight
W_1—vehicle weight at engine start
W_2—vehicle weight at engine shutdown
ϵ—nozzle expansion ratio
Ψ—thrust-to-weight ratio
η—electrical efficiency
ν—heat transfer efficiency

SOME USEFUL PROPULSION EQUATIONS

1. Thrust of rocket: $F = \frac{\dot{W}}{g} v_e + A_e (P_e - P_o)$
2. Expansion ratio: $\epsilon = \frac{A_e}{A_t}$
3. Measured specific impulse: $I_{sp} = \frac{F}{\dot{W}_p}$
4. Mass ratio: $MR = \frac{W_1}{W_2} = \frac{\text{Engine start weight}}{\text{Engine stop weight}}$
5. Overall mass ratio: $MR = (MR_1) \times (MR_2) \times MR_3) \times \ldots$
6. Thrust-to-weight ratio: $\Psi = \frac{F}{W}$
7. Lift-off acceleration: $a = (\Psi - 1)$ g's
8. Ideal velocity change: $\Delta v_i = I_{sp}\, g\, 1nMR$
9. $\Delta v_a = \Delta v_i - \Delta_l + v_r$
10. Theoretical specific impulse:

$$I_{sp} = 9.797\sqrt{\left(\frac{k}{k-1}\right)\left(\frac{T_c}{m}\right)\left[1 - \left(\frac{P_e}{P_c}\right)^{\frac{k-1}{k}}\right]}$$

11. Mixture ratio: $r = \frac{\dot{W}_o}{\dot{W}_f}$
12. Density impulse: $I_d = (I_{sp}) \times (SG)$

13. Total impulse: $I_t = (F) \times (t) = (I_{sp})\ (W_p)$

14. Total Δv (multistage vehicle): $\Delta v_t = \Delta v_1 + \Delta v_2 + \Delta v_3 + \ldots$

15. Thrust of hydrogen-fueled nuclear rocket:

$$F = 6.94 \sqrt{\dot{W}\,[947\ \nu\ Q - 3.76\ \dot{W}\ (T_e - T_i)]} \qquad \text{(See Appendix E)}$$

16. Thrust of electric rocket: $F = 38.4 \sqrt{\eta p\ \dot{M}}$ (See Appendix E)

CHAPTER 4

SPACE VEHICLE ELECTRICAL POWER

THE SUCCESSFUL operation of flight vehicles in space has been made possible, in part, by the availability of electrical power. In general, a space vehicle has neither utility nor purpose without some form of electrical power which is the lifeblood of a space vehicle, vital to the needs of both passengers and equipment. These needs include power requirements for communications, data gathering and handling, vehicle attitude control, guidance and control, life support equipment, and many other areas. With few exceptions, all space vehicles without electrical power become useless debris. The space vehicle electrical power subsystem is, therefore, one of the most important spacecraft elements. Its design often dictates the active functions which a space vehicle can perform. Furthermore, in the future some space vehicles will use electrical power to energize electrical space propulsion systems.

In this chapter, three areas pertaining to space vehicle electrical power are discussed: electrical power generation in space, energy sources and associated energy conversion devices, and the evaluation and comparison of specific power subsystems.

PRODUCING POWER IN THE SPACE ENVIRONMENT

The space environment consists of a near vacuum (10^{-12} Torr at 7000 nautical miles) containing both charged and uncharged particles; electric, magnetic, and gravitational fields; and electromagnetic radiation from both galactic and extra-galactic sources. These and other environmental factors find the space power design and systems engineer working with vastly tighter constraints than the conventional power engineer. Some of the problems of producing electrical power in space are discussed below.

1. Removal of waste heat. A critical and ever-present problem in space vehicles is the disposal of waste heat produced by power generation and power conditioning equipment. Earth based systems for generating electrical power can use convection, conduction, and radiation to remove excess heat. However, the virtual absence of matter around a space vehicle makes radiation the only economical means of waste heat disposal. But, space is an inefficient heat sink; hence, large radiators must sometimes be employed, imposing severe weight and volume penalties. Besides these penalties, a large radiator increases the possibility of meteoroid damage.

2. Meteoroid bombardment. A possible vexing problem is potential catastrophic equipment failure resulting from puncture by meteoroids moving at relative velocities as high as 150,000 miles per hour. Although the probability of this occurrence is statistically quite negligible, the potential damage or erosion effects of meteoroid bombardment must be considered.

3. Very high vacuum. Space systems which must operate in a very high vacuum are exposed to problems of cold welding, outgassing, and leakage. Furthermore, rotating equipment requires lubricants which will not sublimate. Some components require a pressurized environment.

4. Radiation. Among the radiation hazards are cosmic rays, solar proton emissions, and intense ultraviolet and infrared radiation. However, the most serious radiation hazards of space flight to both equipment and man are the geomagnetic trapped Van Allen radiation fields and the solar corpuscular radiation associated with solar flares. This radiation can cause serious damage to solar cells and semiconductors. Shielding presently is considered the most suitable technique for protecting power systems from these radiation hazards.

5. "Weightlessness." In earth-based electrical power generation systems, the force of gravity on a body (weight) is used in separation of vapors from liquids in boiling and condensing fluids, convective heat transfer, and separation of gases from liquids. Since these various processes are used often in generating electrical power, systems cannot use components or processes which operate only with "this end up."

6. Temperature variation. Earth-based systems operate with relatively small temperature variations. In space the temperature of components within the power system may vary through extreme ranges because of solar heating and heat generated by the system itself. The power output of chemical batteries is very sensitive to temperature change. In general, the efficiency of a chemical battery decreases as its temperature decreases, but the output of solar cells increases as their temperature decreases. Also, the useful life of other electrical components depends upon the operating temperature.

7. Launch environment. Earth-based systems for generating electrical power do not operate under adverse conditions. The delicate components of space systems must, however, be designed to withstand the high acceleration and vibration, of the launch phase.

8. Kinetic perturbations. The structure of earth-based systems for generating electrical power is fixed solidly to the earth to counteract any unbalanced forces developed by moving parts. Space vehicles will change attitude due to forces such as those produced by armatures rotating at high speeds or the movement of sun-oriented solar arrays. Counteracting these forces adds to the complexity of the flight vehicle.

Many other problems, such as the limited or nonexistent repair capabilities; the very high premium placed on minimum weight, long life, and high reliability; and the inherent low efficiency of energy devices, combine with the areas already discussed to make the task of electrical power generation in space truly a very difficult one from an engineering design point of view.

TYPES OF SPACE ELECTRICAL POWER SYSTEMS

Electrical power systems for space applications can be categorized by their energy sources into electro-chemical, solar, or nuclear. They then may be categorized further into static or dynamic energy conversion systems. Static systems of current interest include batteries, fuel cells, and photovoltaic, thermoelectric and thermionic devices. Dynamic energy conversion systems consist of either turbine or reciprocating engines coupled to electrical generators. The engines would operate on thermal energy derived from either the sun, a nuclear reactor or isotopes, or the combustion of chemical fuels. Each of these systems will now be considered in greater detail.

Electro-Chemical Systems

The electro-chemical power systems are represented by batteries, fuel cells, and chemical-dynamic systems. Of these, the batteries have until recently played the most important role in furnishing space power. Virtually every space vehicle launched has contained one or more batteries. The reason for their use is that they are very reliable. Furthermore, they are cheap, simple, and, most importantly, available.

BATTERIES.—There exist two types of batteries: primary cells and secondary batteries. Primary batteries are generally considered as one-cycle batteries since they are not recharged after discharge. In this type of battery, chemical energy is converted directly to electrical energy. The greatest utility of the primary battery as a source of electrical power is either in launch vehicles or in space vehicles which have very short missions. This type of battery is also useful for special applications such as providing pyrotechnic power to energize explosive bolts to separate vehicle stages.

The primary battery used most frequently in space applications and having the highest energy density of batteries in common use is the silver-zinc battery. It consists of silver-silver oxide and zinc electrodes immersed in a potassium hydroxide electrolyte. Its useful energy density is in the range of 20 to 100 watt-hours per pound, depending upon the discharge rate. This means that approximately 10 to 50 pounds of batteries must be provided for each kilowatt-hour required. Research is being done on advanced primary cells using more energetic anode materials, such as lithium, magnesium, and aluminum which may provide energy densities in the neighborhood of 150 watt-hours per pound.

However, primary batteries have a limited life, determined by the active composition materials. In general, primary batteries are used extensively for space electrical power requirements of a few milliwatts to approximately one kilowatt for periods less than a week. For example, one hundred and fifty

pounds of silver-zinc primary batteries were used to supply power to the Mercury spacecraft in 1962 and 1963. Because the time in orbit of these craft was not very long, the use of primary batteries was adequate. Similarly, silver-zinc batteries are used to supply power in the lunar module, and in the Apollo command module during the critical reentry period and post-landing period.

Secondary (or storage) batteries are needed for space vehicles which use solar powered energy conversion systems but which require power during solar dark periods. Except for the initial charge, the chemical reaction within secondary batteries provides a means to store the electricity produced by the sun's energy rather than acting as a primary energy source. The secondary battery used most widely in space applications in the nickel-cadmium battery. The advantages of this type battery in comparison to other secondary cells are its demonstrated superior reliability over large numbers of duty cycles, small voltage excursions, high charge rate acceptances, low temperature operation, and long shelf life. However, one disadvantage of the nickel-cadmium battery is its low energy density. Typical energy densities are from 2 to 15 watt-hours per pound, depending upon the required lifetime and charging rate. Thus, for applications which do not require long lifetimes, either silver-cadmium or silver-zinc batteries often are used because of their higher energy densities.

FUEL CELLS.—In recent years considerable research has been expended in developing space qualified electro-chemical fuel cells. Unlike batteries, fuel cells use chemical fuels and oxidants which are stored externally to the cell. The principal parts of one common type of fuel cell are illustrated in Figure 1. Two sintered porous nickel electrodes are immersed in a solution of sodium hydroxide or potassium hydroxide, and there is an external supply of hydrogen and oxygen gases. The hydrogen and oxygen, at a pressure of about 50 at-

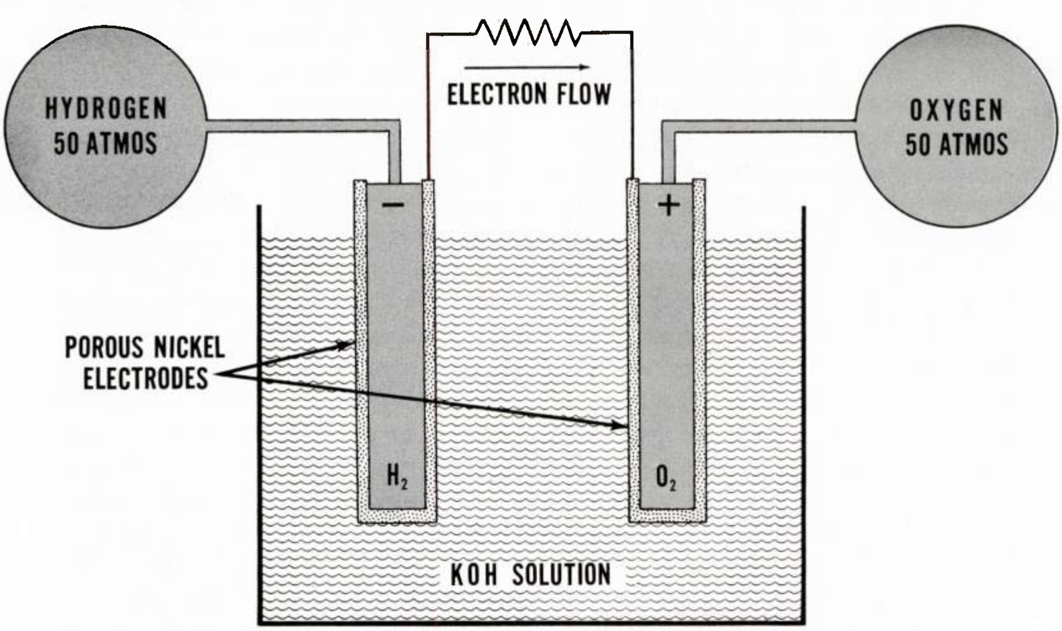


Figure 1. Schematic drawing of a hydrogen-oxygen fuel cell.

mospheres, diffuse through the porous electrodes of the fuel cell. Chemical reactions between the hydrogen and the solution, and the oxygen and the solution, take place on the electrodes. Ions migrate through the solution, and electrons flow through the external circuit.

The chemical reactants must be supplied continuously when the fuel cell is operated, and their reaction product must be removed. The reaction product of the hydrogen-oxygen fuel cell is potable water. Both the reactant consumption and the water produced is approximately 1 lb/KW-hr. Because of its high efficiency, and production of potable water, the hydrogen-oxygen fuel cell is considered particularly suitable for short-duration manned missions. The Apollo program uses hydrogen-oxygen fuel cells as a primary source of space power and drinking water.

Generally, energy densities or specific weight figures for fuel cells are somewhat meaningless because their magnitude varies considerably according to the operating time. This is because the weight of any given fuel cell is practically invariant to the time of operation. On the other hand, the chemical reactant weight is directly proportional to the required operating time and electrical load. From a weight standpoint, fuel cells are highly competitive for applications requiring approximately 10 kilowatts or less of power for operating periods of a few days to three months. Power densities of 200 to 900 watt-hours per pound are within current fuel cell technology for such operational times.

Fuel cells can operate either as an "open cycle" system or as a regenerative, closed-cycle, system. The open cycle system can yield high current but only for the time that the supply of hydrogen and oxygen lasts. Increased operating time necessitates increased supply of reactants with an accompanying increase in the size and weight of the tanks. In a regenerative system energy from some other source, such as solar cells, could be used to produce hydrogen and oxygen from the water formed in the fuel cells. Such a system, if perfected, could have an operational lifetime of several years. However, only the open cycle system has been space qualified so far.

Chemical Dynamic.—Chemical dynamic systems convert the energy available from a chemically fueled combustion engine into electrical power by mechanical means. Such a system falls at the lower end of the power level-duration selection curve (see Fig. 6) because of its relatively high rate of fuel consumption which is two to three times that of fuel cells. However, this type of system generates A.C. power at useful voltage and frequency, simplifying power conditioning requirements. For high power systems, approximately 10 Kw and higher, it tends to be lighter, more rugged, and considerably less temperature sensitive than electro-chemical systems.

Although chemically fueled combustion engines have been extensively exploited for ground and aeronautical applications, ballistic or space efforts have been limited primarily to systems development. The propellants most frequently considered are hydrogen-oxygen, hydrazine, and the broader category of solid propellants.

It had been planned to use chemically fueled combustion engines in both the Navaho and the Dynasoar programs. Although both of these programs have been

terminated, many space power engineers still believe that the chemical engine will have a weight advantage over the fuel cell for operational time requirements of less than two weeks.

Solar Powered Systems

One result of the fusion inferno taking place in the sun is the release of an enormous amount of energy. A considerable amount of this energy occupies the electromagnetic radiation spectrum for approximately ¼ micron to 3 microns (or millionths of a meter). A small fraction of this energy lies in the invisible ultraviolet region; about half of it is the visible light; and the rest is infrared, which accounts for the sun's heat.

At a distance of one astronomical unit (AU), the average amount of solar energy, the "solar constant," is approximately 130 watts/ft^2 of surface aligned at right angles to the sun's radiation. At other than one AU, the energy available is equal to the inverse square of the distance in AU (between the sun and the point of interest) times the solar constant at one AU. Thus, in the vicinity of Saturn, almost 10 AU from the sun, the available solar energy is only approximately 1.30 watts/ft^2, which is $^1/_{100}$ of the available energy just above the earth's surface.

There are basically two primary conversion concepts which have been considered seriously as a means of converting solar energy to space electrical power: photovoltaic and thermal. Of these, the photovoltaic concept has received the principal attention, and it presently provides the basis for the most common source of electrical power for unmanned spacecraft.

Photovoltaic systems.—The photovoltaic effect is the generation of a voltage by photons (the smallest quanta of light) incident on a properly treated surface which lies near the junction of two layers of somewhat different material. Several types of photovoltaic materials can be used to provide useful space electrical power, but because of efficiency reasons, only doped single-crystal silicon solar cells have been used operationally. Typically, contemporary solar cells are thin wafers, either 1 x 2 or 2 x 2 centimeters in area by about 8 to 14 mils thick (without filters). Electrically, the cell is n on p construction (top contact negative) with a base resistivity of 10 ohm-centimeters. The contacts are evaporated silver and titanium. The voltage potential produced by each cell is approximately 0.5 volt, and cells are connected in series to supply the desired voltage. These subunits are then connected in parallel to supply the desired current to be drawn from the complete unit. The solar cell subsystem can be mounted either on panels, paddles, or directly on the skin (body-mounted) of the space vehicle.

One of the inherent weaknesses of the solar cell has been its susceptibility to loss of effectiveness due to nuclear and other radiation in space. In order to decrease this damage, each solar cell is covered with a layer of quartz or sapphire, which absorbs some of the harmful radiation. A reflective coating also is applied to each cell to cause light of wavelengths below approximately 4000 angstrom units to be reflected. The energy in this region has a small effect on the cell, and by reflecting it, the cells can operate at a slightly lower tempera-

ture, thus giving more power. An anti-reflective (SiO_2) coating is also applied to the cell to reduce the loss from reflection at the desired wavelengths. In addition to increasing cell efficiency this anti-reflective coating provides the cell its characteristic blue color. The filter, reflective and anti-reflective coatings, and adhesives add another 4 to 8 mil thickness, yielding a typical total solar cell thickness of 12 to 22 mils.

Solar power subsystems have been used more extensively than any other power subsystem for unmanned space missions. One reason is that solar units do not have to carry any fuel as a source of energy, since the energy source is the sun. However, the production of any appreciable amount of power requires large surface areas. The present level of solar cell technology is 11 to 12 percent efficiency (at air mass zero) and a specific weight of approximately 10 watts per pound. Thus, for example, after accounting for all losses and degradations, a fully oriented silicon solar cell array can provide a specific power of approximately 10 watts/ft^2. A list of solar array parameters for a selected number of spacecraft using typical solar cell mountings is shown in Table I. Note that the body-mounted and fixed paddle solar cell subsystems have specific weights and powers considerably less than those of oriented solar arrays. This is because the maximum electrical power at any location is generated only when the solar cells are aligned directly to face the sun. The solar power available at 1 AU when the solar cells are not aligned is $P = 130 \cos i$ watts/ft^2, where "i" is the angle of incidence (the angle between the direction of the sun's radiation and the perpendicular to the solar cell surface). Furthermore, in the absence of the sunlight, there is no production of electricity.

This latter aspect is an inherent weakness in the use of solar cells in that a solar array subsystem must be supplemented by storage batteries when used on earth-orbiting spacecraft which require continuous power. This is because such spacecraft generally pass into the earth's shadow at regular intervals. The percent

TABLE I

Selected Typical Spacecraft Solar Array Parameters

Spacecraft	*Launch Year*	*Total PWR (Watts)*	*Type Subsystem*[a]	*1x2 cm Cells/ Spacecraft*	*Watts/*[b] *Pound*	*Watts/ Sq Ft*
Pioneer V	1960	25	Fixed Paddles	4,800	—	2.42
Explorer 12	1961	20	Fixed Paddles	5,600	1.8	1.3
Relay I	1962		Body Mounted	8,400	1.4	2.0
Vela	1963	90	Body Mounted	13,236	2.67	2.26
Ranger 7	1964	226	Oriented	9,792	5.5	9.3
OGO A	1964	710	Oriented	32,256	6.79	8.94
Mariner-Mars	1965	680	Oriented	28,224	9.6	9.7
Pioneer VI	1965	82	Body Mounted	10,368	5.47	3.73
IDCSP	1966	46	Body Mounted	7,808	1.96	2.04
OGO D	1967	745	Oriented	32,256	7.12	9.38
Intelsat III	1968	167	Body Mounted	10,720[c]	5.20	3.56
Nimbus B	1969	470	Oriented	11,000[c]	6.0	9.8

[a] The oriented systems, as used in this list, are considered to be any solar array system with some orientation capability.
[b] Including substrate (the base material upon which the silicon wafer is mounted).
[c] 2 x 2 cm cells.

of time spent in the earth's shadow can be as high as approximately 40% for low altitude orbits. Storage batteries are required to provide power during the dark periods and are recharged when the spacecraft is in the sunlight.

Although the fully oriented solar panels have size and weight advantages, these are not obtained without associated disadvantages. Primary among the latter are the difficulties that the fully oriented solar arrays incur in their integration with the spacecraft. Since the spacecraft attitude control subsystem often is used to aim the antennas, separate attitude positioning and a sun sensor are required to orient the solar arrays. Furthermore, to allow positioning and power transfer, the movable solar arrays require such components as bearings, slip rings, and rotary transformers. Another integration problem is to eliminate or minimize the shadowing of portions of the solar array by other spacecraft projecting components (sensors and antennas) and vice versa. The deployment and orienting equipment also cause lower reliability.

In addition, for low earth orbits, atmospheric drag and other perturbations exact a spacecraft stabilization fuel penalty for using any solar system other than a body mounted subsystem. The sum of all of these penalties is such that the use of an oriented system at altitudes less than approximately 175 miles generally is not considered. Furthermore, at higher altitudes, oriented solar arrays generally are used only when large amounts of power (more than 500 watts) are required and/or vehicle configuration and attitude control modes prevent acceptable performance from other systems.

With increasing demands for solar cell generated power in spacecraft applications, various attempts are being made to increase the power to weight ratio and decrease the cost per unit power output. Most of these attempts have been associated with thin-film solar cells. These cells are produced by evaporating a thin film of semiconductor materials, such as cadmium-sulfide or cadmium-telluride, on to a metallized plastic flexible substrate. Although the efficiency of these cells is only 3% to 6%, they still offer potential means of reducing weight and cost of solar arrays in addition to having improved radiation resistance. Another type of solar cell which is receiving consideration is the development of large area dendritic silicon solar cells. Single-crystal silicon cells, 2 x 6 cm in size, are also being studied. The larger areas permit major savings in the manufacture, handling and installation of solar arrays.

Emphasis also is being placed on improved packaging and deployment techniques. Proposed configurations include rigid lightweight biconvex solar array panels and various unfurlable and retractable systems. Several proposed systems function much like a window shade, which can be mechanically unfurled after the spacecraft has been placed in its operational orbit or trajectory. This type of array could also be retractable for periods of powered flight such as during course correction and docking maneuvers.

Projected power densities using these various improvements are approximately 30 to 50 watts/lb, which is almost an order of magnitude improvement of current systems (See Table I). Solar cell proponents, who once thought their horizon was limited to power levels of less than one kilowatt, are now talking of arrays that can supply tens of kilowatts.

SOLAR THERMAL SYSTEMS.—Another possible way to utilize solar radiation is to develop solar heated systems using turboelectric, thermoelectric, or thermionic devices to convert focused solar thermal energy into electricity. The heart of the solar thermal systems is the solar mirror concentrator. Five foot paraboloidal mirrors have been made, for example, which can focus over 85 percent of the incident solar radiation through an aperture of 0.5 inch diameter. This would provide approximately two kilowatts of heat on a thermionic converter at 2000° K. Increased power can be obtained by using solar concentrators that are composed of many petal-like segments or that expand like umbrellas. This construction enables diameters up to approximately 50 feet with correspondingly higher power output.

Although estimates indicate that weight and cost advantages over other power systems may be possible, this is yet to be verified. In recent years, all major focused solar thermal programs either have been cancelled or have been cut back severely. However, some interest still remains, since this type system would be more resistant to charged particle radiation and could operate closer to the sun than solar cell arrays. Conventional silicon solar cells do not function properly at temperatures in excess of 150° C.

Nuclear Systems

The *A*tomic *E*nergy *C*ommission (AEC) is developing nuclear electric power systems under the SNAP (*S*ystems for *N*uclear *A*uxiliary *P*ower) program. The objective of this program is to develop compact, lightweight, and reliable nuclear energy and electrical power conversion devices for space, sea and land uses. The AEC usually starts development for space applications at the request of other government agencies. Since the Air Force was interested in long-life, lightweight, electrical power generation systems for use in space, it requested that the AEC develop such systems. The AEC initiated the SNAP program in 1955. Now the program serves both the Air Force and NASA.

Under the SNAP program, the AEC is developing radioisotope and nuclear reactor systems. The odd-numbered SNAP, such as SNAP-3, use a radioisotope as an energy source, and the even-numbered SNAP, such as SNAP-8, use a nuclear reactor as an energy source. Some of the radioisotope and reactor systems developed for space applications are described below.

RADIOISOTOPE SYSTEMS.—The radioisotope concept for utilizing nuclear energy to provide space power is based on the phenomenon of radioactivity, the spontaneous decay of unstable atoms. This radioactive decay is accompanied by nuclei emission of electrically charged alpha particles, beta particles, and gamma rays. The alpha particles are positive and are the same as the nucleus of the normal helium atom, two protons and two neutrons. The beta particles are negative and are identical to electrons. The designations alpha and beta are terms left over from the period when it had been determined that there were radioactive emanations, but their atomic identity had not been established. Gamma rays, which often accompany radioactive changes, are high energy penetrating electromagnetic radiation. Like high energy x-rays, gamma rays can be very deleterious to biological tissues. In large quantities, gamma rays will also damage structural materials.

The rate of the radioactive decay process is expressed by the half-life of the particular radioactive material. It is a characteristic of each radioisotope species and is independent of the quantity of material or the time at which the decay measurements are started. Thus, the half-life of a radioisotope is that time at which its activity has decreased to one-half of its original value beginning with any given starting time.

Radioisotopes which are sealed in a container generate heat when the container absorbs the radiation emitted by the radioisotope. Part of the heat can be converted into electricity by a suitable conversion system. Radioisotopes are a reliable source of heat, but they are dangerous to handle. If they emit gamma radiation, they require shielding. The lifetime designed into the power system depends on the half-life of the radioisotope. Compensation for the decay of the radioisotope can be made by providing an excess of radioactive material at the start, and just enough at the end, of the operating period. In the space environment, the excess heat generated during the early life of the system must be rejected by radiation. Figure 2 shows the variation in power output for polonium-210 (Po-210) with a 138-day half-life, curium-242 (Cm-242) with a 162-day half-life, cerium-144 (Ce-144) with a 285-day half-life, and promethium-147 (Pm-147) with a 2.6-year half-life.

Although not shown in Figure 2, the power output of a plutonium-238 (Pu-238) radioisotope system would be virtually a constant for the indicated 600 days. This is because the half-life of Pu-238 is 86.4 years.

The Atomic Energy Commission started the SNAP radioisotope power programs several years before Sputnik I. Although the radioisotope power programs suffered a setback in 1959 with the cancellation of an overambitious SNAP-1

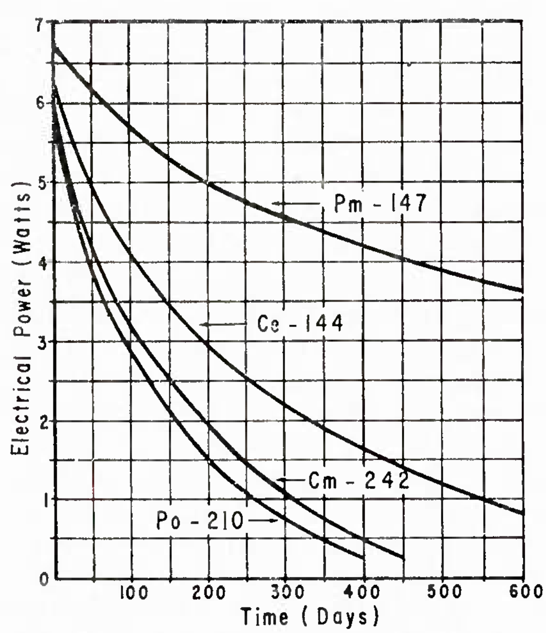


Figure 2. Isotope electrical power versus time.

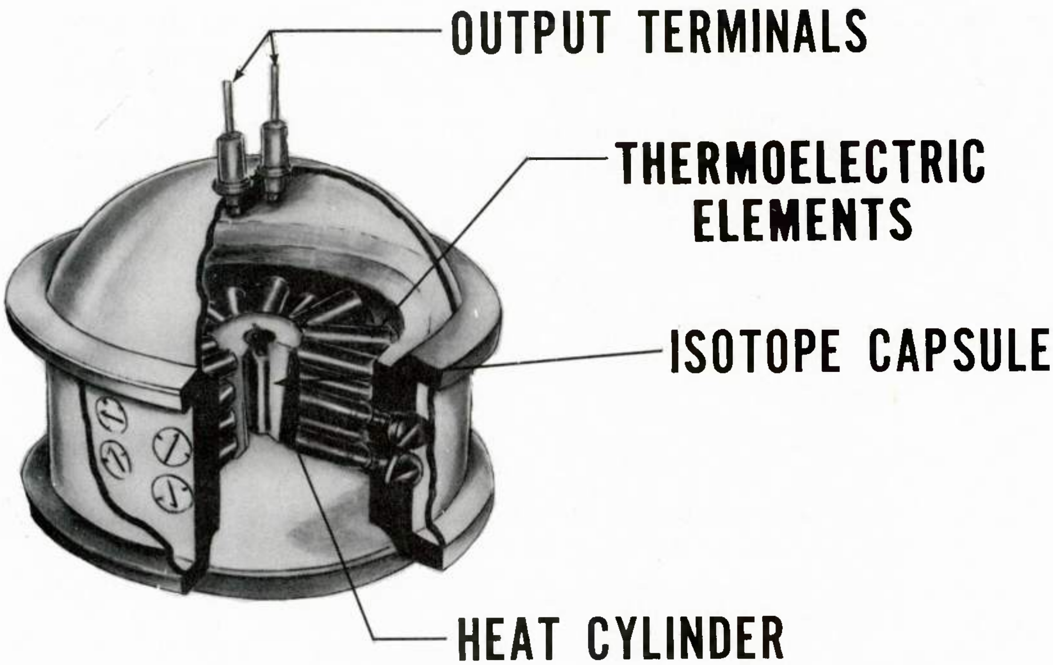


Figure 3. SNAP-3 cutaway diagram.

and SNAP-1A program, isotopic power had developed to the point where it could not be neglected as a potential space power source. For example, the first proof of the practicality of radioisotope power generation took place in January 1959 when the 2.5 watt SNAP-3, a polonium-210 fueled radioisotope generator, was tested and delivered to the AEC.

Polonium-210, which was suitable and readily available, is an alpha emitter. Figure 3 shows a cutaway diagram of this proof-of-principle device which President Eisenhower introduced on 16 January 1959, as the SNAP-3 "atomic battery." It contained 27 lead-telluride thermocouple elements, treated with bismuth to produce negative and positive semiconductors. The thermocouple elements provided parallel paths for the heat to flow from the radioisotope source to the container radiator. These elements were insulated electrically from each other at both the hot (1050°F) and cold (300°F) junctions. The complete SNAP-3 weighed only 4 pounds, and it was only 5.5 inches high and 4.75 inches in diameter.

Because of these efforts, it was possible with only a five-month lead-time to fabricate, fuel, test, and get approval to use a plutonium-238 fueled SNAP-3 radioisotope generator on the Navy's Transit-4-A navigational satellite which was launched in June 1961. This was the first atomic power device used in space. The Pu-238 fueled SNAP-3 was a low powered (2.7 watts), lightweight, and rugged thermoelectric generator. Plutonium-238 was used because of its extremely long half-life (86.4 years), its low gamma radiation, and its high thermal power density per unit mass of fuel. The device weighed about 4.6 pounds and was designed for a five-year life. It provided power for two of the experimental Doppler frequency navigational signal transmitters in the satellite.

The highlights, performance, and status of the Pu-238 fueled SNAP-3 radioisotope generators and additional space radioisotope power systems are shown in Table 2. Note that most radioisotope generators have power densities of approximately one watt per pound. The SNAP-27, a vital cog in the NASA ALSEP (Apollo Lunar Surface Experiment Package), shows the best power density performance, approximately 30 percent higher than the other SNAP devices.

TABLE 2

Space Radioisotope Power Systems

Designation	*Mission*	*Power (Watts)*	*Weight (lbs)*	*Isotope*	*Design Life*	*Status*
SNAP–1	Air Force Satellite	500	600	Ce–144	60 days	Program Canceled in 1958.
SNAP–1A	Air Force Satellite	125	175	Ce–144	1 year	Program Canceled in 1959. One unit successfully ground tested.
SNAP–3	Navy Navigational Satellites	2.7	4.6	Pu–238	5 years	2 in space, 6/61 & 11/61. First in operation. Second failed after 8 months.
SNAP–9A	Navy Navigational Satellites	25	27	Pu–238	6 years	3 launches; 9/63, 12/63, 4/64. 3rd launch failed to achieve orbit. The 12/63 launched SNAP is still operational.
SNAP–11	Moon Probe (Surveyor)	25	30	Cm–242	120 days	Mission cancelled, but one unit successfully ground tested in 1966.
SNAP–17	Communications Satellite	25	28	SR–90	5 years	Design and component test phase completed in 12/64. Presently being held in abeyance.
SNAP–19	Nimbus B	30	30	Pu–238	5 years	Initial launch in 5/68 failed to achieve orbit. Two units successfully launched May 1969. An improved SNAP–19 to be launched in 1971 on a pioneer spacecraft.
SNAP–27	Apollo Lunar Surface Equipment Package	60	46	Pu–238	2 years storage, 1 year Luna operation	One unit successfully deployed in 11/69 by Apollo 12 astronauts. 3 additional units scheduled to be deployed.
SNAP–29	Short Lived DOD and NASA Earth Missions	400	400	Po–210	90 days	Program terminated 6/19.

As far as the future is concerned, it appears that radioisotope generators can be built with power densities of between two and three watts per pound. Although the power outputs of future systems probably will continue to be relatively low (less than 500 watts), higher power outputs are achievable.

REACTOR SYSTEMS.—Nuclear reactor systems are considered to be the most attractive means of obtaining large quantities of space electrical power. A space nuclear power supply consists of three major subsystems: a nuclear reactor heat source and associated radiation shield; an energy conversion subsystem; and a heat rejection subsystem. Such a system derives its energy from the controlled fissioning of uranium-235 in the core of the reactor. The fission energy produced in the reactor in the form of heat then can be converted to electrical power by using one of three approaches: turboelectric, thermoelectric, or thermionic conversion.

As part of the initial SNAP program, AEC started the development of space reactor systems for generating electrical power by asking industry to bid on the development of a system for an Air Force advanced satellite. In the spring of

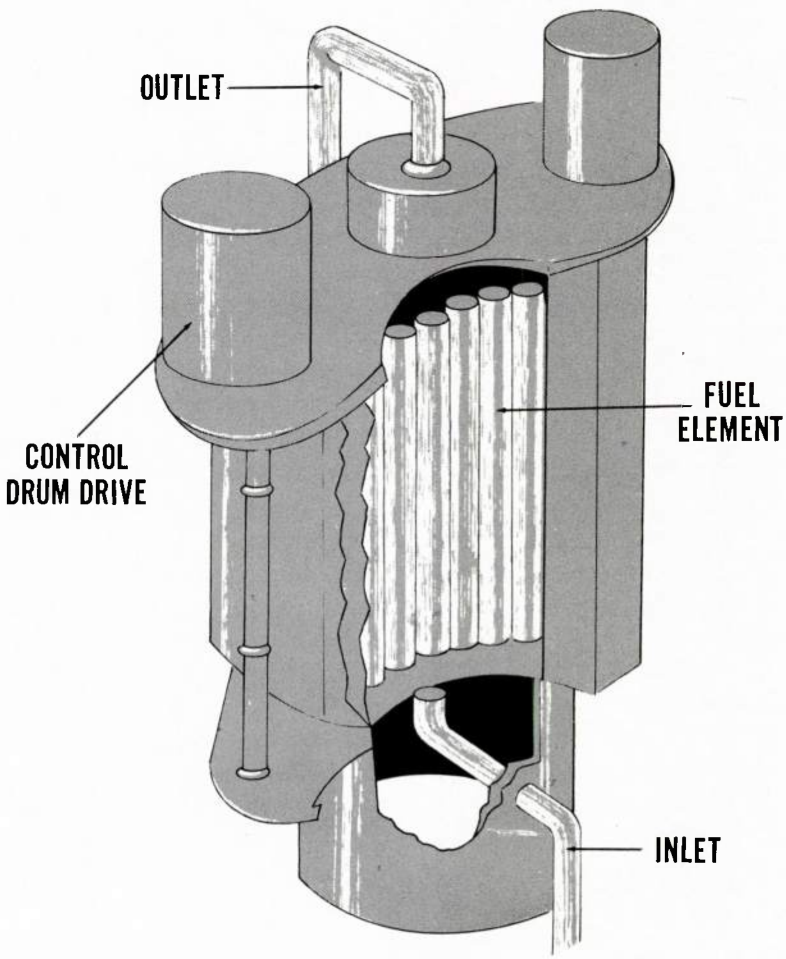


Figure 4. SNAP-2 reactor.

1956, Atomies International's concept, using a mercury-Rankine eyele, was seleeted and designated SNAP-2 (Fig. 4). The objective of this program was to develop a nuclear turboelectric unit generating three kilowatts for use in space. The SNAP-2 reactor uses 37 cylindrieal fuel rods consisting of uranium-235 intimately mixed with zirconium hydride whieh serves as a moderator to slow down the neutrons which maintain the fission ehain.

A sodium-potassium alloy, in the liquid metal state, transfers the heat to the conversion equipment and cools the reactor by circulating through the reaetor core and a mercury boiler-heat exchanger. The eoolant enters the eore at a temperature of approximately 1000°F. and leaves at about 1200°F. The exchanger transfers heat from the liquid sodium-potassium alloy to liquid mereury, the working fluid, converting the mercury into superheated vapor. The mercury vapor expands through a turbine to drive an electric generator, so that electrieity is produeed. To eomplete the eyele, the mereury vapor leaving the turbine condenses to liquid mereury in a radiator before it is returned to the boiler.

The estimated weight of the SNAP-2, when installed in a spaee vehiele as a "nose cone module" and ineluding sufficient shielding for protecting transistors for one year, is 900 pounds.

Several experimental and development reaetors were designed, eonstrueted and operated. Approximately 500,000 Kw_t-hr (thermal power) were produeed from 1959 through 1962 without failure. In 1963, the SNAP-2 program was terminated as a system, but the AEC continued a base program to develop and demonstrate the potential of the uranium-zireonium hydride fuel technology and to improve the power eapabilities of the mercury-Rankine eonversion system.

The first reaetor powered electrieal system to be flight tested in spaee was the SNAP-10A. The thermal energy of the SNAP-10A was derived from a power output modified SNAP-2 reaetor. In the SNAP-2, the reaetor produees about 50 Kw_t at a 1200°F outlet temperature. The SNAP-10 reaetor produees about 30 Kw_t at an outlet temperature of 990°F and an inlet temperature of 880° F.

Furthermore, instead of using a turboelectric heat eonversion system, the heat from the SNAP-10A nuelear reactor was converted direetly into eleetrieity using a silieon-germanium thermoelectric conversion device whieh formed an integral part of the radiator (See Figure 5).

The SNAP-10A was launehed in April 1965 into an earth orbit by an Atlas-Agena vehiele. The system was designed to provide 500 watts of eleetrieal power at 28 vde for a period of one year in a space environment. The power system obtained full power operation approximately 12 hours after liftoff and operated successfully and eontinuously at full power for 43 days. From telemetry data it was established that the probable cause of the shutdown was due to a voltage regulator failure. This resulted in high voltage which caused other components to fail. During the 43 days of sueeessful operation, the SNAP-10A produced 500,000 watt-hours of electrieity, and it met all of its objeetives with the exeeption of the enduranee demonstration.

The SNAP-8 program, consisting of a SNAP-8 reactor and its 35 Kw_e (electric power) mereury-Rankine power conversion system, is a joint NASA-AEC de-

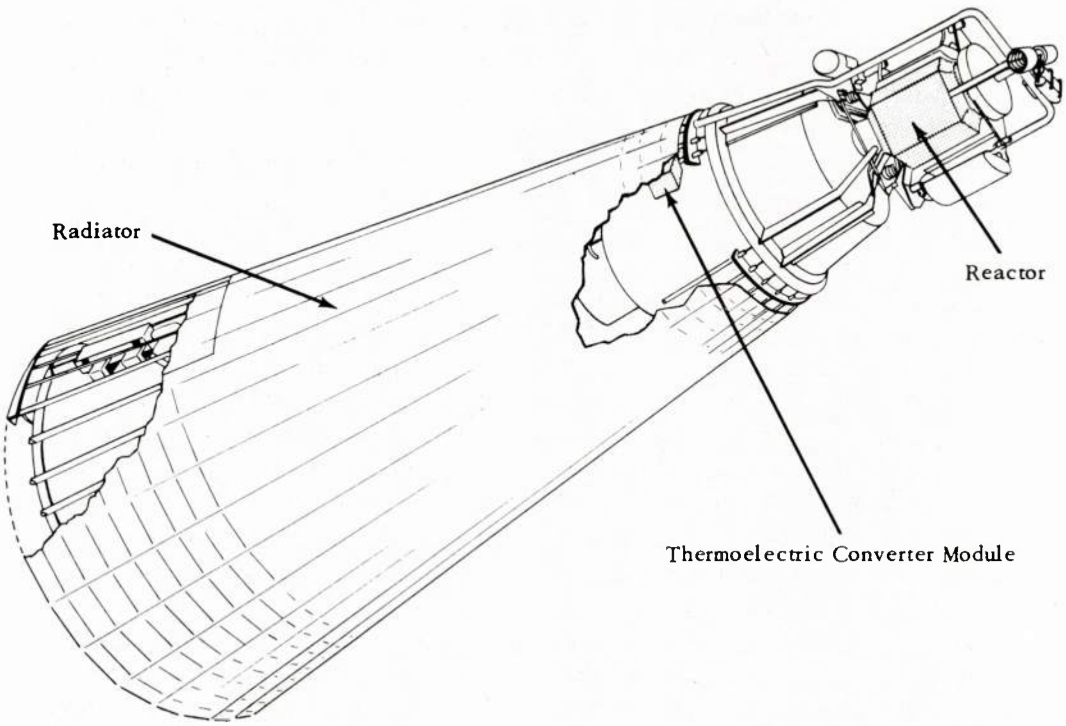


Figure 5. SNAP-10A system.

velopment. The program presently is planned through the ground development phase which is to be completed by 1970-71. The AEC is responsible for the reactor design, and NASA is responsible for developing power conversion equipment and for overall system integration.

The SNAP-8 reactor, designed to produce 600 Kw_t at an outlet temperature of 1300°F, is a scaled-up version of the SNAP-10A reactor. System weight, including the power converter but exclusive of shield and radiator, is approximately 5000 lbs. System weight for manned applications (including shield and radiator) may be as high as 30,000 lbs. NASA's planned use of the SNAP-8 is for applications such as a lunar based electric power plant, on-board electric power plants for large manned laboratories, electrical propulsion for interplanetary missions, and for satellites and probes needing the long life and high power levels which a system powered by a nuclear reactor can best meet.

The Air Force, NASA, and the AEC are all interested in systems with higher electrical power levels than the SNAP-8. Some of the electrical propulsion systems will need megawatts of electrical power for operating periods in excess of one year. However, the development of such systems is a very difficult and costly undertaking, and no well-defined requirements which would justify development of a specific system exist. The SNAP-50/SPUR project, a large reactor power Rankine system aimed at developing 300 to 1000 Kw_e, was cancelled in 1965 and is now relegated to component development status. However, advanced technology development work to provide high temperature liquid metal-cooled reactors for a power system in the multi-megawatt class is still under

investigation at the Lawrence Radiation Laboratory, Livermore, California. The present conceptual reactor designs are based on 3000° F. outlet temperatures and liquid metal working fluid temperatures of 2400° F.

Safety

No discussion of the use of nuclear electric power systems in space flight vehicles would be complete without considering potential radiological hazards. These devices could conceivably present hazards during prelaunch, during powered flight to orbit or to escape velocity, during orbit, or during reentry. Therefore, the following three basic safety objectives must be considered in the design of these devices: (1) Under the worst conditions, these devices should not materially increase the general atmospheric background radiation; (2) At the launch pad, harmful radiation should not extend beyond the device itself or the normal exclusion area; and (3) The device should not produce a local hazard upon return to earth.

Safety has been the basis of many design decisions in the development of SNAP. Compromises in the design of a system have been necessary to achieve a suitable balance between safety and operational characteristics such as weight, simplicity, and reliability. The design decisions have resulted in encapsulated radioisotope heat sources which can withstand the very high overpressures and temperatures which could occur if there were an explosion or fire on the launch pad. Reactor systems are being designed so that full power reactor operation does not begin until the device has been placed safely in orbit. Also, reactor systems and components have been designed to insure burnup at high altitude during reentry. Radiological hazards should not significantly limit the use of nuclear electric power systems in space, when these hazards are anticipated and appropriate measures are taken in design, in establishing handling procedures, and in fixing limits for operation.

SELECTION OF POWER SYSTEMS

The selection of a power system is primarily a function of the space mission and its power requirements in magnitude and duration. However, there is a host of other factors including economic consideration, size and weight, orbital altitude, reliability, and requirements for attitude control. All of these factors make the selection and design of a power system for each specific space mission a custom-built affair.

The information depicted in Figure 6 shows the general areas of application or superiority of the various space electrical power generation systems discussed in this chapter. However, these are only approximate zones of applicability and are based primarily on weight considerations.

As can be discerned from the information depicted in Figure 6, as power requirements rise higher and higher, nuclear reactor power becomes quite attractive. The higher the power requirement and the longer the mission, the greater the probable degree of reactor advantage. However, since a nuclear reactor system has to be of a certain minimum weight before it will produce any power at all,

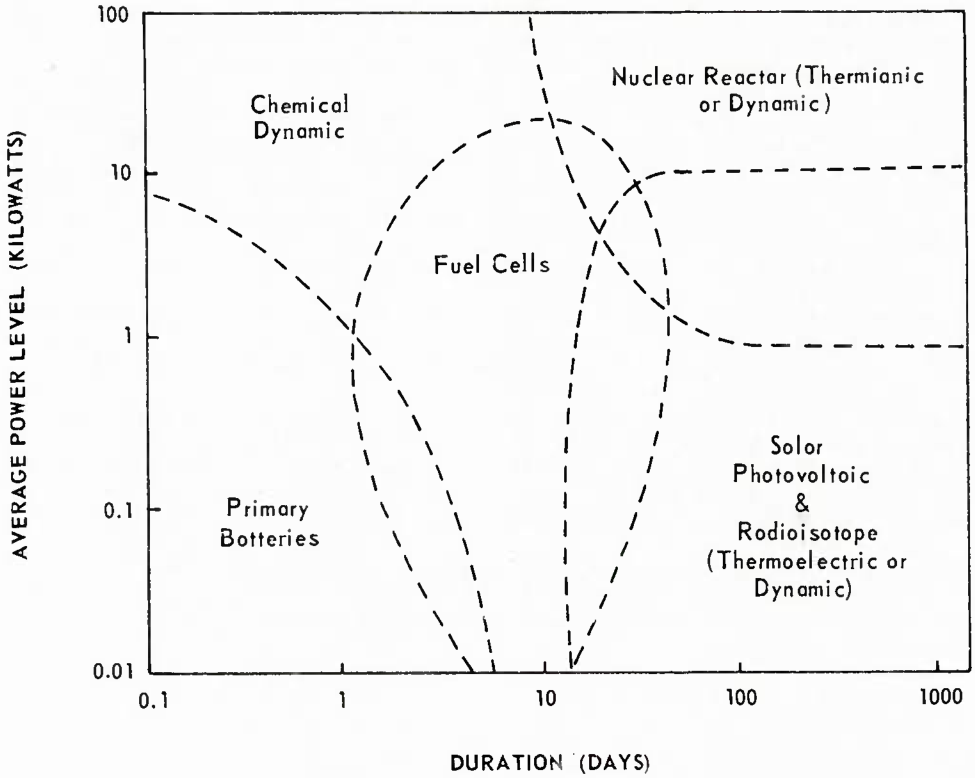


Figure 6. Zones of applicability of space electrical power systems.

solar cells or radioisotope power systems are more competitive at lower power levels. Although radioisotope systems operate in the same power levels and mission duration times as solar cells, their potential seems to be limited to special missions. Their major advantage over solar cells is for missions which require radiation hardening for military purposes or whose orbits traverse radiation belts of high activity, lunar landing operations where they can operate through the 14-day lunar night, and deep space probes to the outer planets where the solar energy is reduced inversely proportional to the square of the distance from the sun. Batteries and fuel cells will continue to dominate the short duration power requirements.

BIBLIOGRAPHY

ARS Space Power Systems Conference, Santa Monica, Cal., 1960. *Energy Conversion for Space Power,* ed. Nathan W. Snyder. New York: Academic Press, 1961. (Progress in Astronautics and Rocketry, Vol. 3)

ARS Space Power Systems Conference, Santa Monica, Cal., 1961. *Space Power Systems,* ed. Nathan W. Snyder. New York: Academic Press, 1961. (Progress in Astronautics and Rocketry, Vol. 4)

American Institute of Aeronautics and Astronautics Aerospace Power Systems Conference. 3d Philadelphia, 1964. *Space Power Systems Engineering,* ed. George C. Szego and J. Edward Taylor. New York: Academic Press, 1966. (Progress in Astronautics and Aeronautics, Vol. 16)

American Rocket Society Space Power Systems Conference, Santa Monica, Cal., 1962. *Power Systems for Space Flight,* ed. Morris A. Zipkin, and Russell N. Edwards (Progress in Astronautics and Aeronautics, Vol. 11). New York: Academic Press, 1963.

AUSTIN, L. G., ed. *Fuel Cells, A Review of Government-sponsored Research 1950–1964* (NASA SP–120). Washington: Scientific and Technical Information Division, Office of Technology Utilization, National Aeronautics and Space Administration, 1967.

BORETZ, J. E. "Power Systems," *Space/Aeronautics,* R & D Issue 1968, 31 July 1968, pp. 85–90.

CROSSLEY, P. A., et al. *Review and Evaluation of Past Solar Cell Development Efforts* (NASA-CR–86979) Princeton, N.J.: David Sarnoff Research Center, Astro-Electronics Applied Research Lab, 1967. (NASA-CR–86979)

FRANCIS, HOWARD T. *Space Batteries* (NASA SP–5004). National Aeronautics and Space Administration, 1964.

GLASSTONE, SAMUEL. *Sourcebook on the Space Sciences.* Princeton, N.J.: Van Nostrand, 1965.

GYLFE, J. D. and Wimmer, R. E. *Reactor-Thermoelectric Power Systems for Unmanned Satellite Applications.* Technical Paper, Intersociety Energy Conversion Conference, Los Angeles, Cal. 26–28 Sep 1966.

HALOEY, D. S., Jr. *The Coming Age of Solar Energy.* New York: Harper and Row, 1963.

HANCHETT, JAMES. "Whither Now SNAP 8," *Nucleonics,* XXV, 5 (May 1967), 54–57.

HOUGH, W. W. *The Effect of Launch Time on the Performance of a Solar Array/Battery Electrical Power System* (NASA-CR–88404). Washington: Bellcomm, Inc., 1967.

MACKAY, DONALD B. *Design of Space Power Plants.* Englewood Cliffs, N.J.: Prentice-Hall, 1963.

Space Power Systems Advanced Technology Conference, Cleveland, Ohio, 1966. *Space Power Systems Advanced Technology Conference* (NASA SP–131). Washington: Scientific and Technical Division, National Aeronautics and Space Administration, 1966.

CHAPTER 5

GUIDANCE AND CONTROL

AS A SPACE VEHICLE begins its flight within the atmosphere, as it coasts through space, and as it reenters the atmosphere, its flight path is monitored and directed by some kind of guidance and control. Any vehicle, traveling through any medium, must be actively guided along its path of movement in order that it may be accurately steered to the destination. Unmanned spacecraft must be directed to their destination either by self-contained on board mechanical and electronic devices or with additional inputs from external stations. Manned spacecraft, as on ships and aircraft, have human navigators. Their task may be either to actively guide and control the vehicle or to simply monitor those functions as they would be accomplished on an unmanned mission.

The function of guidance is separate and distinct from that of control (navigation vs. steering). The guidance function determines the relationship of the space vehicle to the optimum trajectory and then generates a corrective signal when the vehicle deviates from this trajectory. The control function accepts the corrective signal and steers the vehicle so that it will ultimately reach its destination. In a conventional aircraft the two functions are divided between the human pilot and an autopilot. The pilot, who represents the guidance system, determines the deviation from the desired course and initiates corrective signals to the autopilot. The autopilot, or control system, consists of servo controls and airfoil control surfaces which fly the aircraft in response to pilot commands. Although guidance and control are separate functions, they often use either identical or similar components. Gyroscopes, computers, and some electronic devices are often shared by the guidance and control function.

There are two primary differences between guidance of cruise vehicles within the atmosphere and guidance of spacecraft. First, space guidance is a true three-dimensional problem, whereas cruise vehicles operate over a known surface, such as the earth, with a fixed or independently determined third dimension (altitude). Second, space vehicles, unlike aerodynamic cruise vehicles, are accelerated rapidly to a very high velocity and then coast, or free-fall, through a major portion of flight. Therefore, the space guidance system must accurately adjust the velocity and position during acceleration. Small residual errors at thrust termination can develop into large errors later in flight. For example, a ballistic missile is guided during the first several hundred miles of flight and then free-falls thousands of miles to its target. For a 5500 NM range, an error of 1 ft/sec in speed at engine cut-off means an error of about 1 NM at the target.

A guidance and control system consists of three basic elements: sensors, a computer, and a control system. Sensors are either inertial or electromagnetic. The inertial sensing devices are gyroscopes and accelerometers which measure attitude and acceleration. Electromagnetic devices use radio, radar, infrared, and optical instruments which measure range, range rate, and angles. Coordinate conversion, comparisons, and corrective calculations are accomplished by processing data from the sensors through the computer. The space flight *control system* is a closed loop system (servo-mechanism) which uses either aerodynamic forces or rocket thrust, or both, as control forces. In the atmosphere, displacing an airfoil creates lift or drag and steers the vehicle. Both in the atmosphere and in space, deflection of rocket thrust about the center of gravity of a vehicle is used for steering.

Guidance and control for space missions usually is divided into three phases: injection, midcourse, and terminal. The requirements and equipment vary for each phase of guidance and control. The injection phase includes the time from liftoff until the vehicle ascends into an initial orbit. The midcourse phase is primarily a coast phase in which great distance is traversed. The terminal phase can include reentry into the earth's atmosphere or rendezvous in space.

INJECTION PHASE

The objective of the injection guidance phase is the placement of the vehicle at a unique point in space, with a precise velocity, at a certain time. Since the guidance and control system must fly the vehicle along an ascent trajectory, the general shape of the trajectory and the reasons it must take this shape should be understood.

Ascent Trajectory

First, the space booster is launched from a vertical position because it must rise out of the atmosphere as soon as possible to reduce aerodynamic heating and drag. Also, because of its vertical orientation both on the launch pad and in atmospheric flight, it is possible to support the missile axially with a lighter structure. Immediately after liftoff, the control system stabilizes the vehicle in vertical flight. Several seconds later, the vehicle is rolled until the booster keel lies along the desired trajectory plane or azimuth. Then the vehicle begins to pitch slowly so that the velocity vector of the booster will be at the correct flight path angle at burnout. While the booster is in the atmosphere, guidance and control commands are kept to a minimum so that excessive loading on the flexible booster does not occur.

During powered flight, rocket thrust, gravity, lift, and drag forces act upon the space vehicle. During this time large quantities of propellants are expended, and the acceleration of the vehicle increases as it ascends. For a given set of engine and vehicle parameters and mission objectives, a nominal powered trajectory is established by a digital computer simulation some time prior to the launch. Information obtained from the nominal trajectory is set into the guidance equipment for each mission.

The control system must stabilize the aerodynamically unstable vehicle as it flexes and bends. Sloshing of propellants, wind shear, and engine vibration aggravate the control problem.

Systems Used

The two basic types of guidance systems are: radio command and inertial. Each system has unique characteristics which can be advantageous or disadvantageous, depending upon the application.

RADIO COMMAND SYSTEM.—A typical radio command system measures the vehicle position and velocity by means of radar equipment located on the ground. Measured data are fed into a ground-based computer which generates steering commands and an engine cutoff signal. These commands are transmitted by radio command link to the flight control system of the vehicle. A typical radio command system consists of a tracking radar on the ground, an airborne transponder, a Doppler rate-measuring system on the ground, ground computer, and an airborne flight control system (Figure 1).

Use of the radio system makes it necessary to transform data from the ground-based radar into data referenced to inertial coordinates. The radar antennas must have line-of-sight access to the vehicle. The radio system, therefore, requires an extensive ground environment. It may require several radar tracking stations located at strategic points on the earth. Radar tracking and radar control signals may be degraded by disturbances which affect electromagnetic propagation. A guidance station normally can control only one injcction at a time, but the ground equipment can be used repetitively for any number of launches. The airborne cquipmcnt is comparatively simple, lightweight, small and inexpensive.

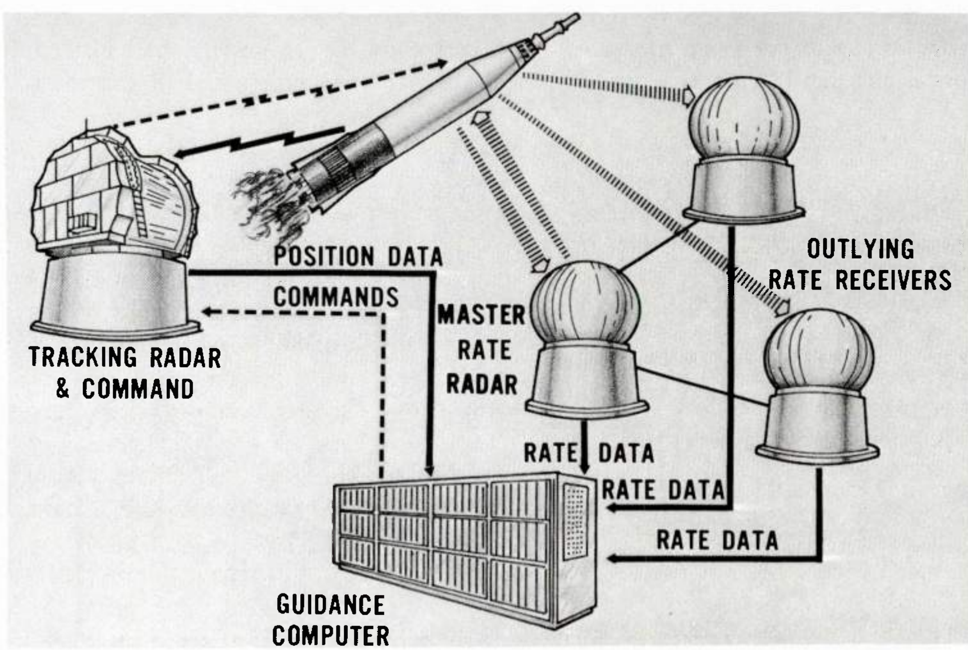


Figure 1. Radio command guidance.

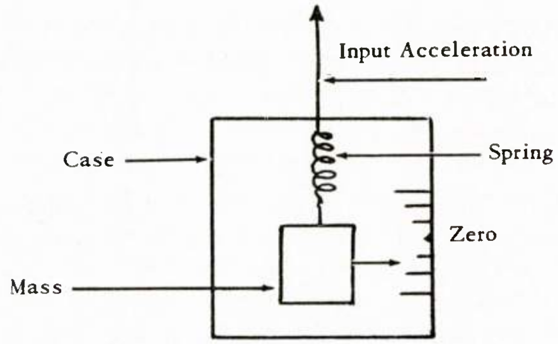


Figure 2. Basic elements of a simple accelerometer.

INERTIAL SYSTEM.—The inertial guidance system measures vehicle acceleration by using three accelerometers mounted on a gyro-stabilized platform in the vehicle. The platform maintains the accelerometer axes in known directions so that acceleration is measured precisely in each of three directions (a coordinate system). The accelerometers may be offset from the navigation reference system and may require coordinate transformations. Figure 2 shows the basic elements of a simple accelerometer. The accelerometer operates on the principle of force application to a mass suspended in a case. As the case accelerates, the mass tends to lag, and the displacement of the mass from the reference point is a measure of the acceleration of the case. The *on-board* computer processes the accelerometer data to obtain velocity and position (Figure 3). Steering commands generated in the computer are sent to the flight control system. All of the equipment described is aboard the vehicle. Launch site location, trajectory parameters, and stable platform orientation must be set into the system prior to launch.

An accelerometer cannot measure the acceleration due to gravity, but, instead, it measures the difference between the true inertial acceleration and the accelera-

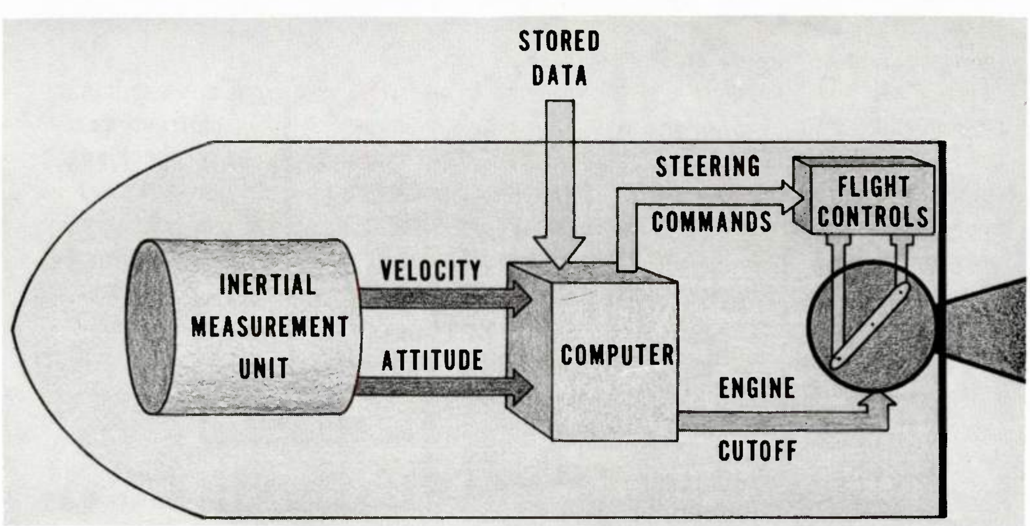


Figure 3. Inertial guidance system.

tion due to gravity. This paradox arises because the acceleration due to gravity is independent of the mass of the body. Therefore, gravity tends to cause the mass, the spring, and the case of the accelerometer in Figure 2 to accelerate at the same rate. Consequently, there is no relative displacement of the mass with respect to the accelerometer case. In other words, an accelerometer only measures the acceleration due to non-gravitational forces, such as thrust, lift and drag.

This inability of the accelerometer to measure the acceleration due to gravity is the fundamental limitation of inertial guidance systems for space vehicle guidance. First, the accuracy of the guidance system can be no better than the accuracy with which the gravitational field is known; this accuracy may not be good enough for interplanetary space flight. Second, and more important, any inertial guidance system will have to have a gravity computer to determine g as a function of position, primarily of altitude. For a spherical homogeneous attracting body, g depends only on altitude ($g = \mu/r^2$).

In inertial guidance systems for ballistic missiles and space vehicles, the location of the launch site and the value of g at the launch site are stored in the computer memory. The information received from the accelerometers is combined with these initial data to compute the velocity and position of the vehicle immediately after lift-off. The new computed position is then used to compute a new value of g. This new value of g is then added to the accelerometer reading to determine the new true acceleration. This process is repeated continuously during the flight. Any error in the accelerometer, no matter how small or how momentary, will cause the errors in velocity and position to increase unbounded with time because the error introduced into the computation loop will feed on itself. The error in the accelerometer reading causes an error in the computed velocity and position, which in turn causes an error in the value of g, which is added to the next value of the accelerometer reading. Even if the acceleromter reading is now correct, the wrong value of g results in another wrong computed position and still greater wrong value of g. This error can be kept within acceptable limits for injection guidance because of the short time of powered flight; in the mid-course phase, however, there must be an independent means of periodically updating the inertial guidance system to remove these errors.

Since the inertial system is self-contained in the vehicle, weight and complexity are increased over that for the radio command systems. The fact that the inertial system is self-contained gives it advantages. The inertial system is not affected by any line-of-sight problem and does not require the radiation of electromagnetic energy. The ground environment is relatively simple; the principal requirement usually is optical data for azimuth alignment of the stabilized platform. Inertial systems allow more flexibility for certain military requirements, such as survivability and quick reaction. Since an inertial system is self-contained, several boosters equipped with this system can be launched simultaneously from a central control facility.

MIDCOURSE PHASE

After thrust termination the path that a space vehicle travels is determined by the laws of celestial mechanics. The vehicle is coasting, except for powered in-

flight maneuvers, until it enters the terminal phasc. While the space vehicle is coasting, various forces act upon it. Aerodynamic forces act upon an earth satellite out to several hundred miles above the earth. There may also be interaction between the magnetic field of the earth and the magnetic material and electric fields on the spacecraft. Additionally, the vehicle is affected by gravitational forces of the moon, the planets, and the sun. Further, rotating machinery, pumps for propellants, or movements of crew members may cause reactive disturbances that affect the attitude of the space vehicle.

Some satellites have no midcourse guidance system, but such a system is required for vehicles which are to be recovered following deep-space missions. Since the midcourse phase is the longest in time and distance traveled, errors early in this phase can build up to unacceptable magnitudes near the destination. Therefore, it is important to navigate during this phase and to apply small amounts of corrective thrust as soon as possible. For deep-space missions, excessive thrust is required to correct for errors which may have been accruing over hundreds of thousands of miles of flight.

Attitude Control

Many earth satellites that do not use thrust to change orbit still require attitude control systems. For many satellites, orientation in a specific direction is required for at least some portions of flight. It may, for example, be necessary to keep an antenna, a camera, or an infrared sensor pointed toward the earth. This may require that a certain part of the vehicle always point toward the earth.

Spin stabilization commonly is used when it is necessary only to maintain the orientation of one axis of a satellite. It is relatively simple, lightweight and reliable. Some satellites use gas jets to control the rate of spin; others make use of the earth's magnetic field. The spin rates used vary from 9 rpm in Tiros satellites to 180 rpm in the Telestar communications satellite.

Another passive means of stabilizing the attitude of a satellite is gravity-gradient stabilization. It is based on the principle that a satellite will align its long axis along the local vertical because of the difference in the force of gravity acting on the top and bottom of the satellite. Although this difference in force, or gradient, seems minute, it can be predominant in the region from about 400 NM altitude to synchronous altitude. Since the satellite will oscillate about the local vertical like a pendulum, there must be some type of damper on the satellite to reduce the amplitude of the oscillations. Gravity-gradient stabilization can be used only in circular or near circular orbit. Solar radiation can be used to orient space probes toward the sun by means of a solar vane.

Although these passive attitude control techniques can be used to keep a spacecraft stabilized, there are times, such as making a midcourse maneuver, when it is necessary to change the attitude of the vehicle arbitrarily. In order to make such attitude changes, an active attitude control system is required which can sense the attitude of the vehicle and can then change it to obtain the desired orientation.

Gyroscopes can be used to determine the *attitude* of spacecraft, but such information will not be accurate over a long time. Gyro-stabilized platforms drift with time and must be corrected by some means for long duration space applica-

tion. One way to realign the gyros is to use optical sighting on celestial bodies. If a reference system can be established by optics, inertial guidance would be necessary for only short periods when rocket thrust was being applied for the small corrective maneuvers. Mariner, Lunar Orbiter, and Surveyor have all used a sun tracker and a star tracker to determine attitude. The star that they have all tracked has been Canopus, because it is the first magnitude star that is most nearly perpendicular to the plane of the ecliptic.

Horizon scanners often are used to determine the orientation of a satellite with respect to a planet. These infrared sensors detect the planet's horizon by the discontinuity in the intensity of the radiation from the warm planet and from cold space. Such a horizon scanner can determine the satellite's orientation with respect to the local vertical, but it cannot determine yaw orientation. A gyro operating as a gyrocompass can be used to provide this information. Ion sensors and magnetometers have also been used to determine yaw orientation.

The methods of changing spacecraft attitude can be based on mass expulsion or on momentum exchange. A mass expulsion device uses propellants stored within the vehicle. The most common type is a reaction jet which expels a gas from a nozzle located so that its thrust axis does not pass through the vehicle's center of mass. The reaction jets use either a cold gas such as nitrogen or argon, or a hot gas such as hydrogen peroxide, or the same hypergolic propellants used for thrusting.

Reaction jets are widely used because they are so simple; however, they do have drawbacks for long missions when large amounts of propellants would be required and for vehicles such as orbiting astronomical observatories that require precise attitude control. For these, the momentum exchange devices can be used. Their operation depends on the conservation of angular momentum. Reaction wheels are used in several satellites, including the Nimbus weather satellite and the orbiting astronomical observatory. If a flywheel attached to a spacecraft is accelerated, a torque will be produced on the vehicle which will cause it to rotate in the opposite direction in order to keep the total angular momentum about the flywheel's axis of spin constant. Three such reaction wheels would be required to stabilize the vehicle about the three axes, roll, pitch, and yaw. It is possible to replace the three reaction wheels with a reaction sphere suspended electrostatically or by gas bearings. Control-moment gyros also provide a means of attitude control by means of momentum exchange. Momentum exchange devices cannot be used without using mass expulsion devices also because the amount of momentum stored in the spinning wheel is limited by the speed at which the wheel will rotate. However, momentum exchange devices do provide a means of reducing the amount of propellant required by the reaction jets on missions of long duration.

Position Fixing

Since the accelerometers cannot measure the acceleration due to gravity, they will indicate zero acceleration during free flight when the only acceleration is the acceleration due to gravity. If it were desired to alter the flight path by application of thrust, however, accelerometers would measure the resulting acceleration. Thus,

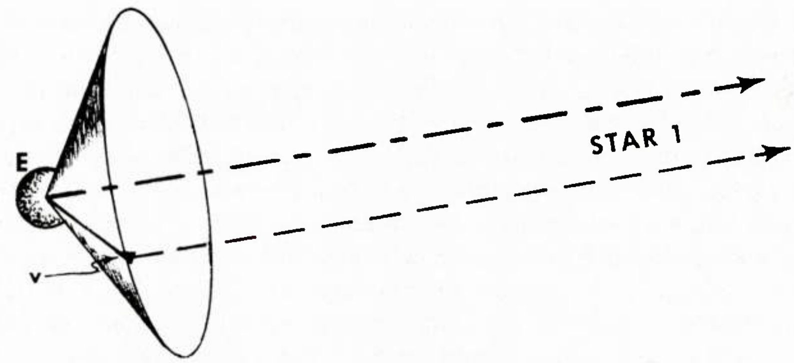


Figure 4. Cone of position.

the measuring ability of accelerometers, and hence, of inertial systems, is limited to flight path *changes* resulting from thrust application during the midcourse phase. Consequently, for missions of long duration, such as lunar or deep-space missions, there must be some other means of determining the free flight trajectory of the spacecraft. This can be accomplished either by ground-based tracking equipment or by optical measurements on board the vehicle.

Tracking of lunar and planetary probes is performed by stations of the Deep Space Net. Because the earth is rotating, three such stations at widely separated locations are required for continuous tracking of the spacecraft. These stations use radar in conjunction with a transponder on board the vehicle to measure range and range rate. Although direction information is also obtained from the orientation of the steerable antenna, it is not accurate enough for trajectory determination. Several more ground tracking stations are required to track satellites in earth orbit. There are seventeen ground stations in the Apollo Manned Space Flight Network. However, for reliability in manned space flights and for certain operational considerations in military space missions, it is necessary to have a

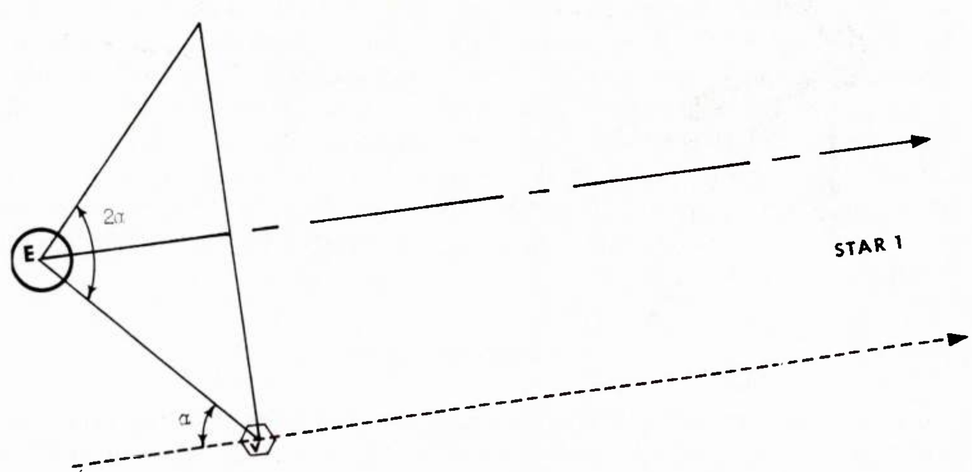


Figure 5. Apex angle of cone of position.

means of determining the vehicle's trajectory that does not depend on the communication link with ground stations.

Optical tracking of earth land marks for orbiting vehicles and optical measurements of celestial bodies for lunar and planetary vehicles offer an on-board means of determining the trajectory of the vehicle. The celestial navigation techniques which marine and aircraft navigators use to locate themselves on the surface of the earth can bc extended to three dimensions for navigation in space. The geometry of a celestial fix in space is described below.

The measurement of the angle subtended at the spacecraft between the line of sight to a near body (the earth) and the line of sight to a star establishes the position of the vchicle on the surface of a cone (Figure 4). The apex of the cone is the center of the earth (or an earth landmark if such can be established). The axis of the cone is in the direction of the line of sight to the star. This axis and the line of sight to the star are considered parallel due to the very great distance to the star. The apex angle (2α) of the cone is twice the supplement (α) of the measured angle between the earth and star (Figure 5).

A second angle measurement between the lines of sight to the earth and a different star establishes a second cone of position (Figure 6).

The cone will have an axis from thc ccnter of the earth in the direction of the sccond star. The apex angle of this cone will bc twice the supplement of the angle measured between the earth and Star 2. These two cones of position intersect in two straight lines meeting at the center of the earth. The vehicle is located on one of these two lines. The ambiguity between the two lines of position could generally be resolved by dead reckoning since one line of position should be close to the intended trajectory while the other line would be at an unreasonable distance from the intended trajectory. A third star measurement with respect to the

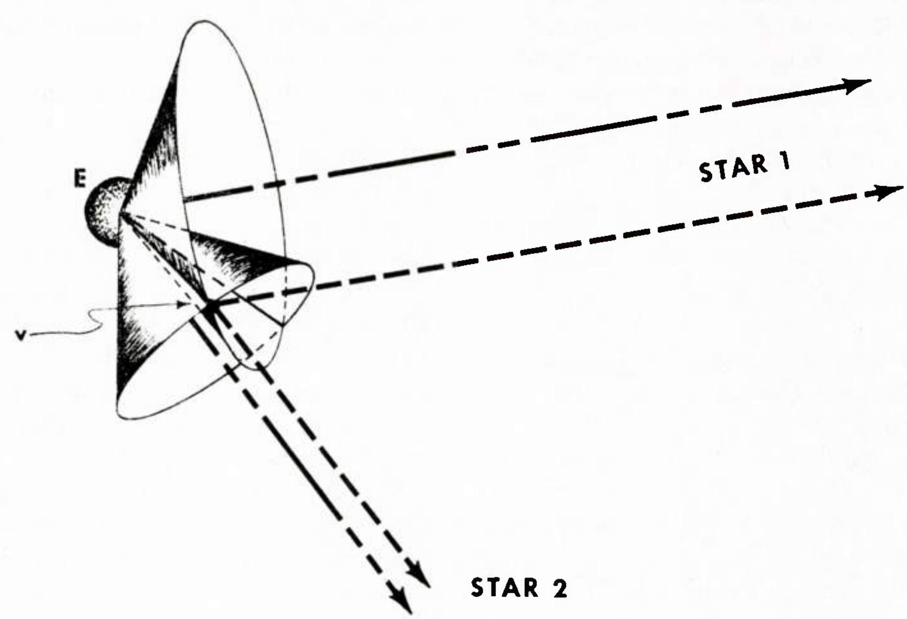


Figure 6. Line of position.

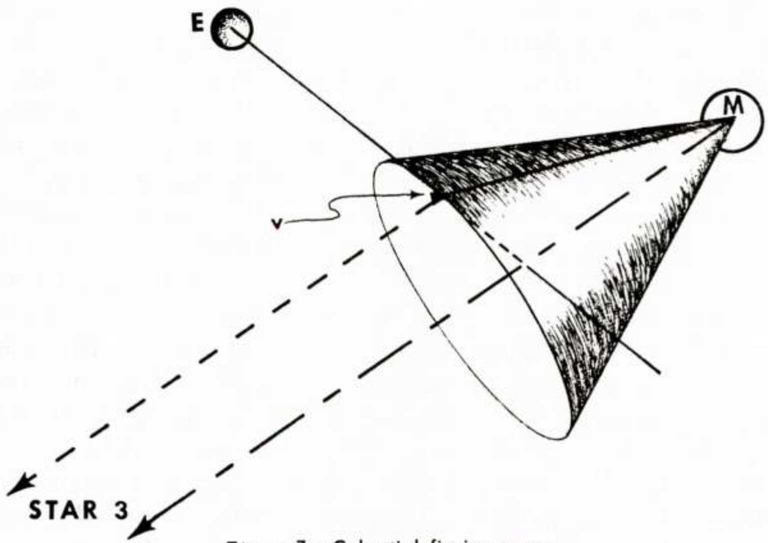


Figure 7.—Celestial fix in space.

earth would also resolve the ambiguity between the two lines of position resulting from the first two star sightings.

However, some method must be used to determine accurately the radial distance of the vehicle along the line of position resulting from the two cones of position. One such method is to employ a third cone of position. Measurement of the angle between lines of sight to a third star and a second near body (the moon) establishes that cone. The intersection of the third cone and the previously established line of position would be at two points (Figure 7). Again, the ambiguity can be resolved either by selecting the point closest to the intended trajectory or by taking another sighting.

There are several disadvantages in applying this celestial technique. First, the algebraic equations used in the analytical solution of the intersection of three cones are non-lincar. Their solution would be difficult even with the assistance of an on-board computer. Second, all sightings have to be taken at the same time. This would be impractical. Third, there is no satisfactory method of incorporating redundant measurements to compensate for instrumentation errors. All of these difficulties can be overcome by modifying the procedure to base the calculations on an assumed position, somewhat as aircraft navigators do. In this case, the assumed position would be that point on the intended trajectory where the vehicle would be if there were no crrors. The angle between the near body and the star that would be observed if the vehicle were at that point is precomputed. The angle between the near body and the star is then measured at the vehicle's actual position. The difference between the measured angle and the computed angle is used to locate the actual position with respect to the assumed position on thc intended trajectory. Three or more such measurements could be taken in a short time and then adjusted to common time to locate the vehicle in three dimensions with respect to the assumed position.

Since mid-course guidance is three dimensional, a three-dimensional coordinate system is required to locate the celestial bodies as well as the vehicle. And since

the planets and the moon are moving in the coordinate system, an ephemeris, or table giving the location of the bodies in this coordinate system at various times, is required.

TERMINAL PHASE

In the terminal phase of flight, the guidance system uses the characteristics of a target as a reference for completing the flight. Typically, for a space vehicle returning to earth, the terminal phase could begin at an altitude of several hundred thousand feet and thousands of miles from the destination. This phase might commence at 50 or 100 NM from the position of another space vehicle for the rendezvous of two vehicles. In each case, the guidance system must complete the mission with minimum use of thrust. Therefore, for reentry through the earth's atmosphere, aerodynamic braking is used to dissipate the energy of the returning vehicle. This, in turn, imposes stringent requirements on the guidance system in preventing the vehicle from encountering excessive heat and deceleration loading.

Rendezvous

For rendezvous of two moving vehicles, it is necessary to match their position and velocity vectors. The terminal gravitational attraction between the vehicles is negligible.

Consider the problem of an earth launch to rendezvous with a satellite. In general, the orbiting and ascent paths would not lie in the same plane. In addition there would be some error in adjusting the position and velocity vectors of the interceptor to those of the satellite. This means that a series of vernier corrections are required to complete rendezvous.

Optical instruments, which are relatively simple and accurate devices, may be useful for tracking some space vehicles. Others may require detection and tracking by radar. Either radar or optics can provide relative velocity and position information. This, coupled with inertial devices for attitude control, makes it possible to control properly the thrust of the vehicle in completing the rendezvous.

Earth Reentry

On reentry to the earth, the vehicle's trajectory, and hence the guidance and control system, will be greatly affected by the terminal attraction of the earth and the earth's atmosphere. As an example, consider the guidance requirements for a manned vehicle returning from the moon, entering on the first pass rather than by successive grazing entries. Consider first a ballistic vehicle and the associated entry corridor, which is the region between the overshoot boundary and the undershoot boundary (Figure 8). The overshoot boundary, as set by lift-to-drag ratio, determines whether the vehicle remains in orbit for successive passes—that is, whether it does or does not reenter. The lower boundary is set by the peak deceleration which is acceptable. For a ballistic vehicle with a peak deceleration of 10 g, this corridor is about 8 NM deep. The corridor can be made

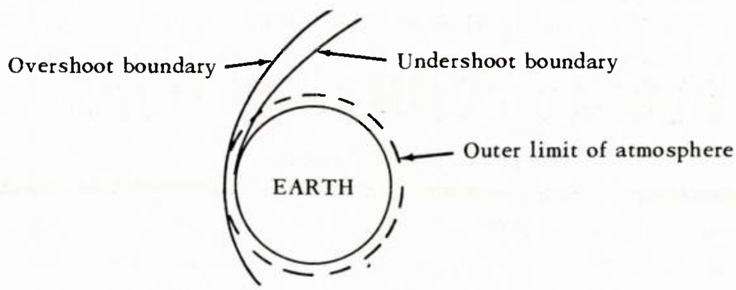


Figure 8. Entry corridor for a space vehicle.

significantly deeper by using a lifting vehicle. This means that the accuracy requirement for guidance can be significantly reduced. For example, with a vehicle having a lift-to-drag ratio of 1, the depth of the corridor can be increased to about 60 NM.

A lifting vehicle allows more flexibility in the reentry maneuver because it permits some option in landing rather than a ballistic free-fall. However, many flight control problems are introduced when the vehicle configuration is changed. It appears that most maneuvering will be accomplished after the critical period of peak heating has been passed. Inertial devices with performance similar to that achieved for ICBM guidance systems are required to maintain the necessary flight path angles during reentry.

BIBLIOGRAPHY

Air Force. Air Force Systems Command. *Space Planners Guide.* III-D, On-Board Space Vehicle Guidance and Navigation; IV-A-4, Attitude Control. Washington: Air Force Systems Command, 1965.

Blasingame, B. P. *Astronautics.* New York: McGraw-Hill, 1964.

Broxmeyer, Charles. *Inertial Navigation Systems.* New York: McGraw-Hill, 1964.

Fernandez, Manuel and Macomber, George R. *Inertial Guidance Engineering,* Englewood Cliffs, N. J.: Prentice-Hall, 1962.

Glasstone, Samuel. *Source Book on the Space Sciences.* Chapter 4, "Guidance, Tracking, and information System." New York: Van Nostrand, 1965.

Hymoff, Edward. *Guidance and Control and Spacecraft.* Holt Library of Science, Series III. New York: Holt, Rinehart, & Winston, Inc., 1966.

Koelle, H. H., ed. *Handbook of Astronautical Engineering.* New York: McGraw-Hill, 1961.

National Aeronautics and Space Administration. *Control, Guidance and Navigation of Spacecraft.* NASA SP–17. Washington: U.S. Government Printing Office, 1962.

North Atlantic Treaty Organization. Advisory Group for Aerospace Research and Development. *Space Navigation, Guidance, and Control,* ed. John Edwin Miller. London: W. and J. Mackay and Co., Ltd., 1966.

O'Donnell, C. F., ed. *Inertial Navigation Analysis and Design.* New York: McGraw-Hill, 1964.

Pittmann, G. R. ed. *Inertial Guidance.* New York: John Wiley and Sons, 1962.

Scull, J. R. *Spacecraft Guidance and Control.* NASA SP–68. (Space Technology, Vol. 4) Washington: Scientific and Technical Information Division, National Aeronautics and Space Administration, 1966.

CHAPTER 6

GLOBAL COMMUNICATIONS

CLOSELY associated with the guidance and control system for a space vehicle is the communications system. Unlike other aspects of the space program, however, global communications essentially presents no new problems. Since the early part of the present century when Marconi developed his "wireless" telegraphy, man has been sending messages through space. Today, the means already exist for communicating as far as the known limits of the solar system. The problems that the space communications engineer is asked to solve are not involved with the generation and propagation of electromagnetic waves, but rather with problems that arise because of the complexity of the system components and the nature of the host vehicle. Such problems are those involved in allowing for the cubage and mass of the equipment in the host vehicle and in generating electrical power in space.

This chapter will acquaint the operational commander with the practical aspects of communications. Effective communication is essential; the commander must be able to exercise instantaneous control, through his communication system.

COMMUNICATION SATELLITES

Satellite systems provide global command and control networks and permit the simultaneous transmission of a greater number and variety of high quality messages. In addition, there is a higher degree of reliable transmission and a greater degree of survivability designed into the system.

In this age of jet aircraft, nuclear power, and moon rockets, present communication facilities simply have become outmoded. These facilities have served well and will continue to do so in the future, but they are not good enough. They are unable to handle the greatly increased volume of communications. Present communications are disrupted easily. A slight increase in solar flare activity can cause radio communication blackout between Europe and North America. A fishing trawler may cut accidentally the North Atlantic telephone cable. While these events present difficult problems to civilian communications, the implications of even temporary communication isolation from the rest of the world are far more serious for a military commander.

During the past several years, communication satellites have attracted considerable interest within both industry and government. There have been successful launchings of communications satellites, and development programs have been initiated both within the civilian community and the Department of Defense.

Rather than discuss specific programs, this chapter will present an underlying philosophy of communication satellites.

Objectives

What should be some of the major objectives in developing a system of communication satellites? One objective is reliable communications which provide uninterrupted service over long periods. The system should have high capacity with capability for handling large volumes of all types of traffic. It should be flexible to serve the maximum number of potential users. There should be minimum delay in transmission. Each of these factors will be discussed in more detail, and the communication satellite will be compared with some existing facilities.

Reliability

Two types of reliability are of interest. The first is propagation reliability. The high-frequency band has always been subject to the vagaries of the ionospheric layers which surround the earth. Thus, only a portion of the HF band actually is useable at any given time over a particular path. In addition, multi-path effects seriously limit the amount of information which can be transmitted over a given channel. Added to these limitations are the blackouts which may result from ionospheric disturbances caused by sunspot activity. High-altitude nuclear explosions can introduce similar disturbances. One is forced to the conclusion that HF radio via the ionosphere is less than satisfactory as a propagation medium.

By contrast, a communication satellite of the active repeater type employing line-of-sight transmission at microwave frequencies would be extremely reliable from a propagation standpoint. However, the communication satellite introduces a second type of reliability problem, that of reliable unattended operation for long periods in orbit. Scientists have demonstrated that a reliable communication satellite can be developed with life times of more than one year if the following practices are used: components of proven reliability are selected; all components are operated well within their ratings; the satellite design provides adequate protection during launch and while in the space environment; and, finally, adequate use is made of redundancy to further increase the probability of successful operation.

High Capacity

Through the years man has been dependent, almost exclusively, on high frequencies in the band 5 to 30 megahertz* for long range global communications. These frequencies are shared among all countries and must support both military and civilian applications. The narrow range of frequencies, and the propagation characteristics discussed previously, seriously limit the total communication capacity.

* Hertz—the unit of frequency, cycles per second. KHz (kilohertz) is 1,000 cycles per second. MHz (megahertz) is 1,000,000 cycles per second.

It is not surprising that great interest continually exists in new techniques which promise to open additional areas of the frequency spectrum to long-range communications. Examples are ionospheric and tropospheric scatter propagation. The communication satellite will open the complete range of frequencies to 10,000 megahertz for long range communications, thus providing nearly 1000 times the spectrum available in the HF band.

Flexibility

One requirement is to provide sufficient flexibility in a system so that new or changing demands around the world can be satisfied without major overhaul or replacement of facilities. A disadvantage of the submarine cable, for instance, is lack of flexibility with a point-to-point, fixed-plant facility. On the other hand, a system of communication satellites in 24-hour equatorial orbits, by providing wide bandwidths and essentially global coverage, places minimum restraint on the number and location of ground stations served and the volume of communication furnished to each.

Minimum Delay

Another objective is to speed up communications. All too frequently, congested facilities cause urgent message delays because propagation conditions are poor. Some of the advantages of communication satellites which were discussed previously, such as the wide bandwidth, will not only make possible worldwide television broadcasting in real time, but also will reduce delays in all types of communications.

SURVEY OF TYPES OF COMMUNICATION SATELLITES

Thorough review of all factors has led to the conclusion that an active repeater satellite in a 24-hour equatorial orbit offers the most promise for advancing global communications. However, in view of the reliability problems and the anticipated costs of a communication satellite program, it would be well to review briefly the pros and cons of alternative approaches which are under development or have been proposed.

Passive Reflector

The passive system uses a reflecting surface which cannot amplify or retransmit signals. Some of the advantages of a passive system are its inherent reliability and the possibility of being shared by a large number of users, operating over a wide range of frequencies. However, a typical system, operating between two locations 2000 miles apart, would require 24 balloons 100 feet in diameter in randomly spaced orbits at 3000 miles altitude for an outage time of 1 percent. Substantially more satellites would be required to provide longer range or wider coverage than the example cited; therefore, such satellites do not appear to be an economical solution to providing truly global communications.

The passive reflector satellites would be more attractive if they could be stabilized in attitude, permitting use of more efficient reflecting surfaces, or if the satellite orbits could be synchronized so as to reduce the number of satellites required. However, the satellites then cease to be "passive" and reliability is no longer inherent in the system.

Active Repeater

Active communications satellites receive signals, translate them in frequency, and amplify and retransmit the signal at a higher power level. They can be considered to be repeater stations in space. The use of active communications satellites makes possible the use of smaller ground terminals, which, in turn, enhances flexibility of military operations. In addition, because of the increased radiated power over that of the passive reflector, the transmission path loss is not as great a problem; therefore, the active satellite can be placed in orbit at much higher altitudes. Because the number of satellites required to provide continuous coverage for a number of ground terminals varies as the orbit altitude changes, the higher the system is placed, the fewer satellites will be required.

Medium-Altitude System

Continuous coverage from medium altitude would require from 18 to 24 satellites in orbit. Even then, there would be a switching problem at the ground terminal as one satellite passed from view and a new one approached to take its place. The ground terminals require steerable antennas as well as computing equipment to calculate the trajectories and furnish look-angles (acquisition data) for antenna orientation.

A major advantage of the medium-altitude system is that orbit injection is simplified and precise position stabilization is not required. The booster requirements are not as great. The medium altitude satellite has the additional advantage of requiring less radiated power and antenna gain than comparable signal reception requirements for higher altitudes.

Synchronous-Altitude System

Three satellites at an altitude of 19,360 NM and equally spaced in 24-hour equatorial orbit, because of the earth's rotation, appear to remain fixed to an observer on the earth. Such a satellite system will provide complete global coverage except for the extreme polar regions. The high altitude of the synchronous orbit (Fig. 1) makes each satellite visible from 40 per cent of the earth's surface. Since the satellite appears to be motionless in the sky, service may begin or continue with only one satellite in orbit and functioning. However, in actual operation, 10 or 12 satellites would be used to provide the desired coverage. While this number is less than that required for a medium altitude system, the booster requirement is much greater, since it is necessary to inject about the same amount of satellite weight at a much higher altitude. Another disadvantage is the time delay which could be experienced in connection with telephoning.

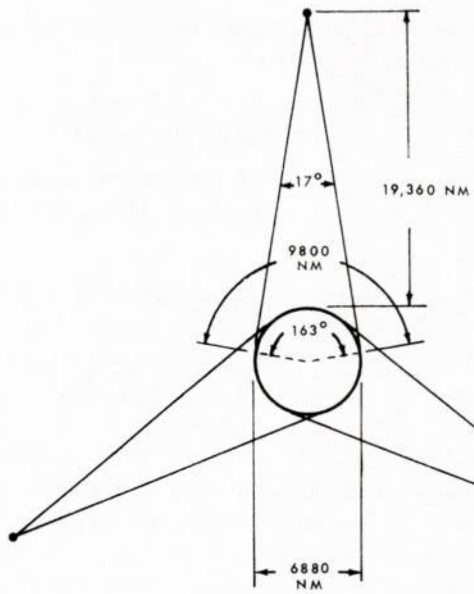


Figure 1. Concept of three communication satellites in 24-hour equatorial orbit, showing geometrical relationships in the equatorial plane.

The two-way propagation delay through a synchronous satellite is approximately 0.6 second. This delay is not objectionable, however, if no echo exists. Proper balancing of equipment and use of echo suppressors are solving this problem. Another major problem is establishing and maintaining a truly synchronous orbit.

Constraints

To appreciate the capabilities of space communications more fully, it is well to consider the natural constraints and design limitations.

1. Natural constraints

a. Line-of-sight. A line-of-sight path must be established between the transmitter and the receiver.

b. Space attenuation. Power radiated from the transmitting antenna is distributed over an ever-expanding portion of a spherical surface. The resulting decrease in power density (power per unit area) reduces the energy "captured" by the receiving antenna and is known as space attenuation.

c. Noise. Noise is introduced at each stage of the communications process. The most significant contributions are from the medium through which the communications are sent and from within the receiver itself. Noise reduces the ability of the receiver to detect weak signals.

2. Design limitations

a. Transmitter power. Power must be available at the transmitter.

b. Receiver sensitivity. This is a measure of the minimum signal strength with which the receiver can be gainfully operated.

c. Receiver noise figure. Noise injected into the system by the receiver itself constitutes a basic limit on the minimum detectable signal.

d. Bandwidth. Bandwidth of the system is limited by many considerations. Most important, the capacity of a system to transmit data is directly proportional to its usable bandwidth.

3. Data processing. Only the transmission of information constitutes a profitable expenditure of energy for a communications system. All data does not constitute information. The effective capacity of a system is directly proportional to the efficiency with which the transmitted data represents information.

4. Modulation. The effectiveness of the communications system varies greatly with the modulation technique used. (Modulation is the process of imposing signal data on a carrier much as the lips and tongue "modulate" the "carrier" generated by the larynx.)

The most important constraints on the communications facility are explained in more detail.

Line-of-Sight Transmission

A prime limitation on the communications facility is the necessity of establishing line-of-sight transmission between the transmitting and receiving antennas. This is always true and applies in every case, but virtual line of sight may be obtained when direct line of sight is not possible.

Direct line-of-sight coverage from a transmitter is shown in Figure 2. The receiver is below the radio horizon represented by the dotted line (located where a

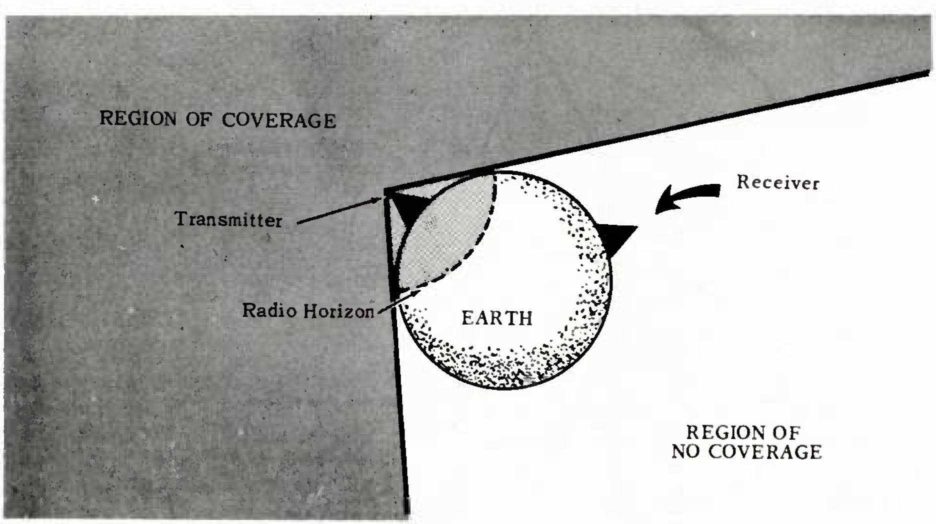


Figure 2. Direct line-of-sight coverage of a transmitter.

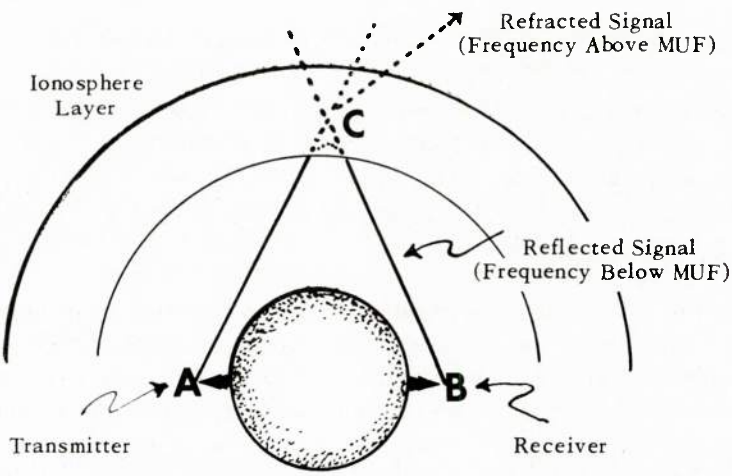


Figure 3. Reflection and refraction of radio waves by the ionosphere.

cone with the apex at the transmitter would be tangent to the earth's surface). The question naturally arises as to the possible extension of the coverage. After all, radio transmissions are received at points far below the radio horizon. The reason this happens is easy to understand. The near-earth environment is not empty space. It is full of oxygen, nitrogen, water vapor, charged particles, magnetic fields, and other material. Of particular interest is the ionosphere which consists of layers of charged particles that have the property of transmission, refraction, or reflection of radio waves, depending on the properties of the layer, the geometry of transmission, and the frequency involved. Below certain critical frequencies, or maximum usable frequencies (MUF), the charged layers of the ionosphere affect radio waves in much the same manner that a partial mirror may reflect and refract light waves (Fig. 3).

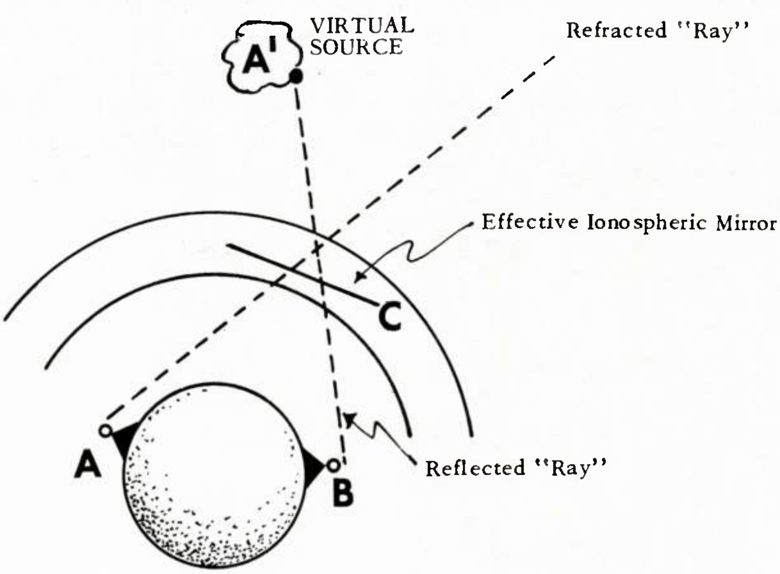


Figure 4. Virtual image of a transmitter established by reflection from the ionosphere.

In Figure 3, the transmission from A is reflected by the ionosphere to the remote position B, which lies below the radio horizon of the transmitter. As shown by Figure 4, the receiver does not actually "see" the transmitter as being located at point A. As far as the receiver at B is concerned, the transmission is originating at point A′, the virtual image of the physical transmitter established by the ionosphere mirror at point C. Line of sight is thus established between a virtual source at A′ and the receiver at B, and the requirement of line-of-sight transmission from the transmitter to the receiver is fulfilled.

This is analogous to seeing around a corner by means of a mirror, as shown in Figure 5. With the mirror arranged as shown, the object is not seen at its actual position, but at a virtual position determined by optical geometry. In using this analogy, remember that, in the case of the ionosphere, the virtual source is not so well defined as is the actual source. The virtual source has become enlarged; its boundaries are blurred by the turbulent condition of the ionosphere.

This "method of mirrors" seems to solve the problem in a simple way. Nature furnished the reflecting medium, so why look for a better scheme? Something better is needed, however. The ionosphere, being in a constant state of flux, is unreliable as a transmission medium. The properties of the ionosphere vary from year to year, from day to day, and even from hour to hour (Fig. 6). Also, energy in the maximum usable frequencies reflected from the ionosphere is too low to provide the desired information-carrying capacity on a channel using three frequencies. (The higher the carrier frequency, the higher the data capacity of the channel.) Use is desired for carrier frequencies of an order of magnitude higher than that permitted by the maximum usable frequencies of ionospheric reflection technique.

Another important problem arises because of the fact that the ionosphere is widely distributed spatially and that it has considerable thickness. Therefore, several transmission paths of varying lengths are established between the transmitter and the receiver. The result of this "multipath transmission" is fading, garbling, and marked reduction in the effectiveness of transmission.

To employ a virtual source to "see around the curve of the earth," the properties of the system must be fixed, so that they are predictable, constant, and localized in space. Theoretically, one way of doing this would be to construct tall antennas, as shown in Figure 7.

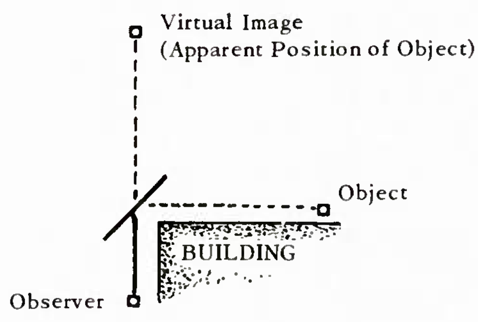


Figure 5. Virtual image of an object formed by reflection in a mirror.

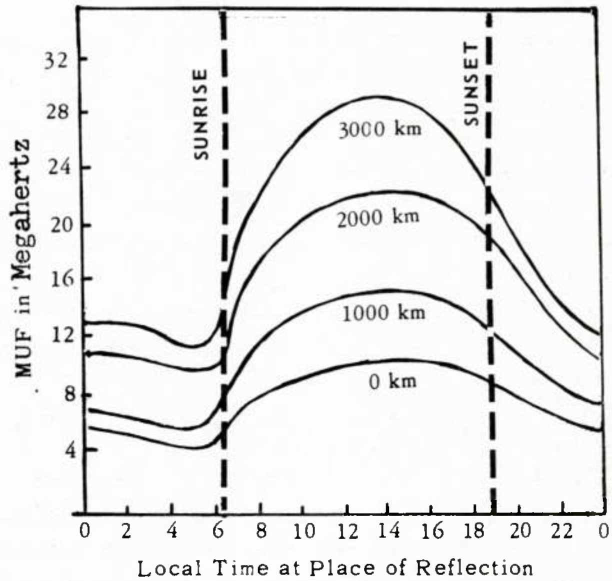


Figure 6. Variations in the ionosphere as reflected by hourly changes in the maximum usable frequency.

Someone calculated, however, that antennas for trans-Atlantic communications would have to be 360 miles high. To construct such towers would take the gross national product of the United States at the current rate for the next 70,000 years.

A more practical way is to place a satellite at the virtual source A′, shown in Figure 4. The satellite serves as a controlled, predictable, space-localized transmitter (or passive reflector) at the virtual source. By selecting frequencies which are essentially independent of the ionosphere, the system will have the capacity, reliability, and coverage to please the most demanding commander.

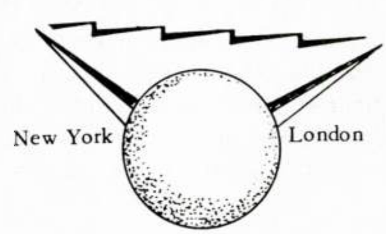


Figure 7. Tall antennas used to overcome the curvature of the earth.

Space Attenuation

When the receiver is far removed from the transmitter, the signal is weaker than when the receiver is in the proximity of the transmitter. One of the reasons this occurs is because of a purely geometrical relationship known as space attenuation. As the term implies, the signal strength decreases as the

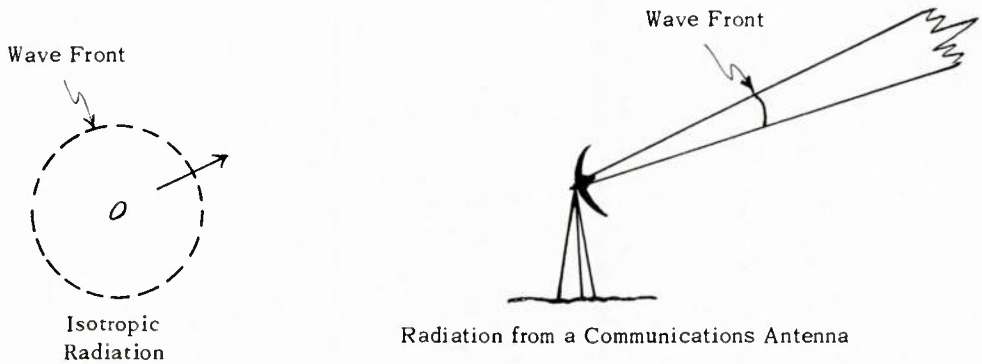


Figure 8. Comparison of isotropic radiation with directional radiation from a communications antenna.

intervening transmission path becomes longer. However, this decrease is not a loss in the sense of an irreversible conversion of electromagnetic energy to some other energy form, such as heat.

Attenuation of the signal constitutes one of the basic limitations on communication over extensive distance. Space attenuation is not a loss. It occurs because the receiving antenna is of limited physical dimensions. If an antenna could be constructed to inclose the source completely, space attenuation would not occur.

Many techniques are employed to minimize the effects of space attenuation. In general, these techniques recognize that isotropic radiation is seldom a necessary specification for a given system. Transmitted power does not have to travel uniformly in all directions from the source. Instead, the power is focused to travel in the direction preferred. An antenna which can focus power is called a directional antenna. The most familiar example of a directional antenna is perhaps the parabolic radar antenna which serves a dual purpose: (1) to form a narrow beam, and (2) to increase greatly the power density within the beam over that accomplished by radiation from an isotropic source. Even in interplanetary communications, isotropic radiation would not be necessary since the orbits of the planets are roughly in a common plane. Thus, transmissions from a point within the solar system would use a pancake-shaped pattern. This would have greatly enhanced capability over an isotropic radiator.

Noise

Most people are familiar with the crackling static that plays havoc with programs on the broadcast band during electrical storms, particularly if the station is not strong. Indeed, during heavy electrical disturbances one may have to stay close to the receiver in order to hear. In this case, the communications range has been severely curtailed by natural interference, i.e., noise injected into the system through the antenna.

Noise also is generated within the receiver itself. If an ordinary radio is tuned to a frequency at which there is no incoming signal, and the volume is

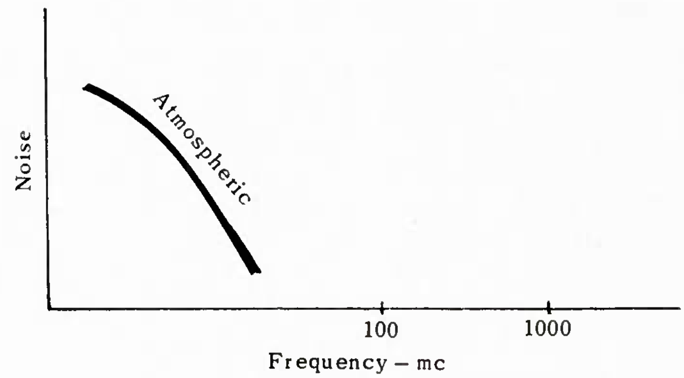


Figure 9. Variation in noise level with the frequency of the radio band.

increased, a hissing background can be heard. If the receiver is powerful, and if little static is present, full volume will nevertheless give a very loud noise output at the speaker. The source of this noise is in the components themselves and cannot be eliminated. By careful design, however, this source of noise can be reduced to a very low level.

Two questions suggest themselves: Which source of noise gives more trouble? How does one measure the noise level of a receiver? First of all, natural interference predominates in the lower part of the radio-frequency spectrum, that is, in the low and high frequency bands or in frequencies up to about 30 MHz. At frequencies above 30 MHz, the noise generated within the receiver itself predominates to the extent that external natural noise can be neglected. During an electrical storm a radio may be unreadable, but television is not affected. This happens because radio operates at frequencies far below 30 MHz (in the region of natural noise), and television operates above 30 MHz where natural interference does not present a major problem. Since the maximum usable frequency for transmission by ionospheric reflection is below 30 MHz, natural interference is present when ionospheric transmission is used.

Modulation

Modulation is the process by which intelligence or signal data is imparted to the radio wave or carrier. A radio signal is not a simple, single-frequency (monochromatic), sinusoidal oscillation which generally is associated with wave motion. Rather, it is a carrier wave varied by some feature, such as amplitude or frequency, which varies with the intelligence (data) being transmitted. This variation of some feature of the radio wave is called modulation. The fundamental radio frequency upon which the modulation is imposed is the carrier. The carrier frequency of a radio signal is the frequency, or number on the dial, to which the receiver is tuned in order to receive a particular signal.

The two most common modulation techniques, amplitude modulation and frequency modulation, might be represented graphically, as shown in Figure 10.

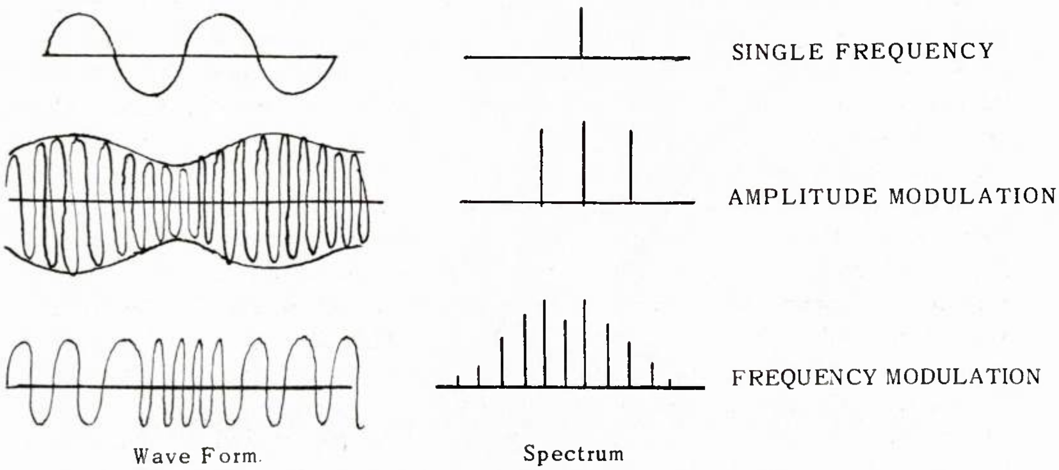


Figure 10. An amplitude modulated radio wave and a frequency modulated radio wave, compared with a single-frequency wave.

Bandwidth

The band of frequencies which the signal occupies is called the signal bandwidth, and the components of the signal above and below the carrier are referred to as the signal side bands. It is apparent that the receiver window, or bandwidth, must be at least as large as the signal itself to receive the entire signal. It is interesting to note that the carrier and side-band components of a radio signal are not a mathematical fiction, but they can actually be observed with proper equipment, such as a spectrum analyzer. The significance of the side-band components lies in the fact that they contain the data or intelligence of the signal, and they must be transmitted by the transmitter, propagated through space, and received and reproduced at the receiver equally well; otherwise, part of the intelligence of the signal will be lost. Consequently, the data rate, or capacity, of a communications system largely determines not only its operational characteristics but also its technical requirements. Examples of bandwidths required for common types of communications service are: 100 hertz for teletype, a low capacity or low data-rate system; 3000 to 5000 hertz for voice communications; and 6,000,000 hertz for commercial television, which is a high capacity or high data-rate system containing both aural and visual detail.

As shown in Figure 10, the modulated radio wave is a sinusoidal wave which has been distorted in either amplitude or frequency, or in some other manner, to provide intelligence. When only the carrier frequency is present, the bandwidth can be theoretically infinitesimally narrow, to encompass only that particular frequency. However, when the carrier is modulated, the bandwidth must encompass the carrier frequency and the frequencies superimposed upon it as side bands. The total number of frequencies represented in the modulated wave is a function of the rate at which the wave is varied or modulated. For example, a radio wave with a carrier frequency of 10 KHz which is modulated by a 1000 hertz tone will contain frequency components from 9 KHz to

11 KHz. Therefore, the bandwidth to pass such a radio wave would of necessity have to be two KHz in width, when considering the normal commercial broadcasting stations. Special types of transmission, such as single side band, do not requirc as large a bandwidth. But, in standard practice the bandwidth must be at least as large as the highest modulating frequency, and in many applications very much larger.

Earlier, it was indicated that the minimum usable signal power level is determined by noise: noise entering the system with the signal at the antenna, and noise generated in the receiver itself. Electrical noise is generally considered to be Gaussian in nature, uniformly distributed throughout the spectrum. Wide bandwidths, then, while desirable from a data-rate and capacity viewpoint, are undesirable from the viewpoint of noise. The wider the bandwidth, the greater is the average noise power that enters the system, both at the antenna and within the electrical circuits. Of course, the power transmitted could be raised to restore the signal to an acceptable minimum signal-to-noise ratio, but power is not always easy to come by. An additional factor, the relation of power to bandwidth, must also be considered. The total power in the signal consists of the power in the carrier frequency and that of the side-band frequencies. In an amplitude modulated signal as much as one-third of the total signal power resides in the side bands, and the entire bandwidth must be transmitted, or part of the intelligence will be lost. Quite obviously, wide bandwidths, which are required for high data rates and high capacity, also require high power to maintain a usable signal-to-noise ratio.

To summarize: high data rates require large bandwidths and high power, and the average noise power is proportional to bandwidth.

Conclusion

Serious shortcomings exist in present global communication facilities. Communication satellites promise greater capacity, wider coverage, more flexibility, and fewer delays than present transmission systems. The operating cost of communication satellites is intimately tied to reliability. With sufficient emphasis on reliability, satellites will assume the prominent role in communications for which they appear destined.

BIBLIOGRAPHY

Congress. House. Committe on Government Operations. *Government Use of Satellite Communications—1968.* Thirty-fourth Report by the Committee on Government Operations. House Report 1836. 90th Cong., 2d Sess. Washington, D.C.: U.S. Government Printing Office, 1968.

Filipowsky, Richard F. and Muehldorf, Eugen I. *Space Communications Systems.* Englewood Cliffs, N.J.: Prentice-Hall, 1965.

Mueller, George Edwin and Spangler, E. R. *Communication Satellites.* New York: John Wiley & Sons, Inc., 1964.

Wall, V. W. "Military Communication Satellites." *Astronautics and Aeronautics,* 6, 4, Apr 68, pp. 52–57.

CHAPTER 7

LASERS

LASER—the word has an unfamiliar sound about it, rather like some science fiction term which has been picked up for everyday use. Stressing the importance of the laser, General Schriever, past Commander of the Air Force Systems Command, has said, "This revolutionary discovery may prove to be more important to the world than the development of the ballistic missile, the discovery of the transistor. . . . The laser will have a profound impact on every scientific and technical discipline."

The laser truly has proven to be one of the most exciting and potentially far-reaching technological developments in recent years. Immediately following Dr. Theodore Maiman's success in constructing the first laser in 1960, laser research advanced at an unprecedented rate. As a consequence, lasers have blossomed in these few short years from laboratory curiosities into systems ranging over almost the entire science and engineering technology fields of endeavor.

Although laser technology is still in its infancy, the current widespread acceptance and future potential of laser devices in both military and nonmilitary applications dictate that forward looking Air Force officers gain an awareness of this growing field. A host of questions immediately arise: What is the laser? How does it work? What are its uses? This chapter will examine each of these questions with the general aim of presenting a clear picture of what the laser is and how laser light differs from ordinary light.

SOME CHARACTERISTICS OF LASERS AND ORDINARY LIGHT

The term *laser* is an acronym. It is taken from the words *L*ight *A*mplification by *S*timulated *E*mission of *R*adiation. The meaning of these terms will become more apparent as the operation and properties of the laser are understood. To facilitate this understanding, some properties of ordinary light will now be reviewed.

Visible Light

First, recall that visible light is a form of electromagnetic radiation (see Figure 1) which carries energy at a velocity of 186,000 miles per second. Light (like radio, radar, and x-rays) is a form of electromagnetic energy which travels in waves of various lengths or frequency. Waves have three characteristics:

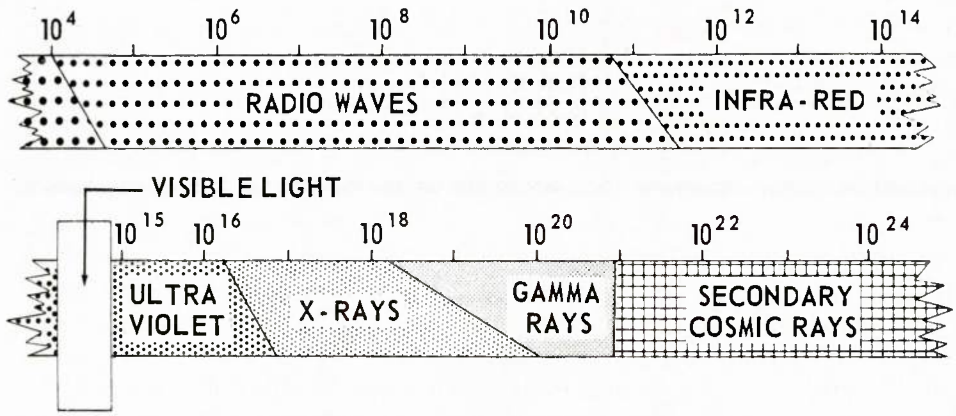


Figure 1. The electromagnetic spectrum, with the frequency of the various forms of electromagnetic energy expressed in Hertz (cycles per second) using power of ten notation.

frequency (number of waves passing a given point each second), wavelength (distance between successive peaks), and phase (instantaneous position of a wave in relation to some position or time reference point). The relationship between wavelength, λ, and frequency, f, is given by

$$\lambda = \frac{c}{f} \tag{1}$$

where c is the speed of light.

The frequency or wavelength of electromagnetic energy determines to a great extent the properties of that energy. For example, the visual sensation of color is a direct function of the frequency of the electromagnetic radiations.

The information depicted in Figure 1 shows that visible light frequencies extend to almost 1,000 million million (10^{15}) Hertz (cycles per second). This frequency is approximately one billion times higher than radio broadcasting frequencies (500 to 1500 kilo Hertz) and at least 100,000 times higher than most radar frequencies. Note that the scale used in Figure 1 is logarithmic. Although the bandwidth of the visible spectrum is shown smaller than that of the radiowave spectrum—it is not. If it were possible to draw the depicted information using a linear scale (where each identical increment of distance anywhere on the chart represents the same number of Hertz), then the width of the visible light spectrum would be almost 4000 times greater than the total radio wave spectrum.

Generation of electromagnetic energy in the optical region is not new. Man has been able to perform this feat since he built his first fire or turned on his first electric light. However, until the advent of the laser, all sources of electromagnetic energy in the optical frequency spectrum (at useable power levels) were neither coherent nor monochromatic. By investigating the properties of ordinary white light emanating from the sun or from an incandescent light bulb, one can discover that such light forms a very complex wave and is not mono-

chromatic—it consists of more than one color. Scientists have shown that ordinary white light contains literally billions of frequencies and is a composite of the entire frequency spectrum of visible light: red, orange, yellow, green, blue, and violet. Furthermore, each frequency component travels in a helter-skelter or random direction and phase relationship with varied amplitudes. The resultant composite complex wave is defined as *incoherent radiation.* This incoherent nature of ordinary light is related directly to the manner in which it is generated.

All light radiation and absorption by materials result in electron activity. From quantum theory, it is known that only a fixed number of electron orbits are possible with each atom. However, electrons can jump from one orbit to the next depending on whether they gain or lose energy. Energy absorption occurs when an atom gains energy from an outside source. Its electrons, particularly those in outer orbits, jump to higher energy orbits farther away from the nucleus. This action raises the internal energy of the atom. The lowest and normal energy level of an atom is known as the ground state, and higher energy levels are known as excited states. Unless constrained, the electrons of excited atoms will, within a very short time, fall back to lower energy levels or directly back to the ground state energy level. In this process, some of the excess energy is radiated as electromagnetic energy.

This electromagnetic energy may be radiated as heat, light or in other forms. Typically, conventional light sources are hot bodies. While atoms of the light sources are continually raised to an excited state, others are falling back to lower energy levels in a random autonomous fashion. In this process, the atoms emit their excess energy in the form of photons (the smallest quanta of light) in an uncoordinated direction, frequency, and time relationship. This emission process is designated as spontaneous emission. The light that emerges is a composite wave of a large number of independent waves that reinforce and cancel each other in a random fashion. This explains why the radiations which comprise ordinary light are incoherent.

The significance of incoherency becomes apparent when one considers the possible use of ordinary light as an energy source for an application such as communications. The radio transmission of intelligence is accomplished by modulating a communications carrier frequency. The modulation may be amplitude modulation (AM), frequency modulation (FM), phase modulation (PM), or a host of other modulation schemes. Reception and interpretation of the impressed intelligence is dependent on the apriori knowledge of the composition of the communications carrier frequency. For example, in a frequency modulated system the intelligence is contained in the modulated frequency deviations about the fixed carrier frequency. However, the instantaneous frequency composition of incoherent ordinary light radiations are completely unpredictable because they contain billions of random frequencies, phases, and amplitudes. Hence, without a known frequency, amplitude, or phase reference, it is an impossible task at the receiver end to demodulate an incoherent signal. About the only thing one can do is transmit data by turning the signal on and off as in semaphore communications.

Another limitation of light prior to the advent of the laser was the difficulty of obtaining high power in a narrow range of wavelengths. Although one could

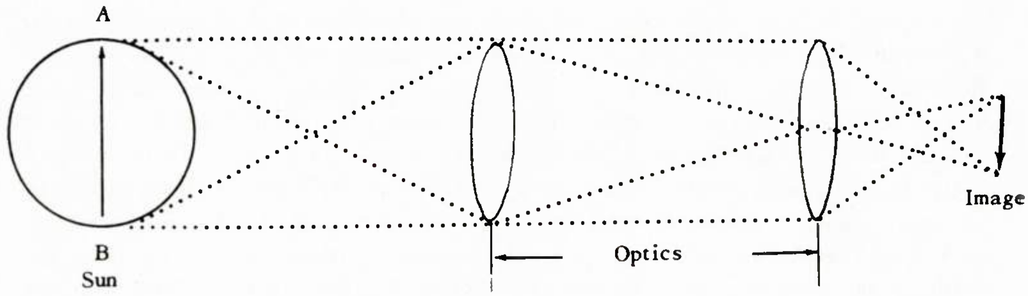


Figure 2. Focusing the energy from a finite source.

filter incoherent light down to a narrow frequency spectrum, it was at the expense of intensity.

However, even without the narrow spectrum limitation, ordinary light generation is at relatively low energy density levels. Thus, consider the energy density which can be achieved from an energy source of finite dimensions, such as the sun. As shown in Figure 2, attempts to focus the sun's rays with an ideal optical system would, at the very best, bring the sun into focus. Irrespective of the optics used, it would be impossible to bring the rays emanating from A to bear on the same point as the rays emanating from B. Also, in the barrel of the optical system, no way exists to bring the rays into exact parallelism. Under this condition of a finite source, the best that can be done is to obtain an energy density at the focus which is equal to the energy density at the source.

The sun's total radiation at all wave lengths is estimated to be seven kilowatts per square centimeter of solar surface. Because of the size of the sun, this represents a large amount of energy. Yet, this solar surface energy density is modest when compared to achievable laser energy densities.

Laser Light

A striking characteristic of a laser is its ability to generate an intense, monochromatic, highly directional, and coherent beam of electromagnetic energy.

The laser energy is considered monochromatic because it is not unusual for the frequency bandwidth to be less than one part in 10 billion of the center output frequency of the laser radiations. Greater chromatic purity than this can be achieved. In addition, the laser emissions are spatially coherent—the laser radiations travel in unison in the same direction and phase.

A prerequisite for this laser action to take place is a pumping source which can cause a population inversion, that is, more atoms in an excited state than in their normal ground state. The laser properties result because atoms which have been pumped into a population inversion state are extremely unstable and readily release their excess energy. In accordance with quantum theory, excited atoms can exist only at discrete energy levels which are separated by ranges of unallowed energies.

By appropriate triggering, excited atoms can be stimulated to fall back to lower energy levels in phase and radiate laser light at a frequency which satisfies the relationship:

$$f = (E_2 - E_1)/h \tag{2}$$

where h is Plank's constant (6.625×10^{-27} erg sec) and E the allowable energy of specific levels. The intensity, coherence, and spectral purity occur because the stimulated emission return to the lower energy level is a resonant process. This action takes place almost instantaneously with light radiated cooperatively in a synchronized chain reaction fashion instead of in the usual independent uncoordinated spontaneous emission mode.

The directionality of the laser radiations, on the other hand, is due both to the coherence property and to the geometry of the laser. Most lasers consist of a column of lasing material sandwiched between two parallel mirrors, a totally reflecting mirror at one end and a partly reflecting mirror at the other end. This arrangement forms an optical cavity resonator. The length of the cavity is a whole number of wavelengths, satisfying the relationships expressed by equations (1) and (2). The stimulated emission is forced to develop along the longitudinal or mirror-axis of the laser. Any photons which do not travel along the mirror-axis either escape through the sides of the laser cavity resonator or are reabsorbed by the lasing material. Those that travel along the mirror-axis bounce back and forth between the mirrors and produce an avalanche of photons which erupt through the partially transparent mirror. The result is a highly collimated (parallel) beam. The synchronized waves depicted in Figure 3 illustrate a two dimensional representation of the resulting monochromatic, coherent electromagnetic radiation. These waves reinforce each other in a cooperative manner.

Whereas the optical focused energy density of ordinary incoherent light is at most equal to the energy density at the source, the same remarks do not apply to the laser radiations. The rays from a laser are highly collimated, and with the use of suitable optics, laser energy can be focused to a very small point. Diffraction limits approaching 10^{-4} radians can be readily achieved. Thus, even for a nominal amount of laser output power, the energy density at the focus

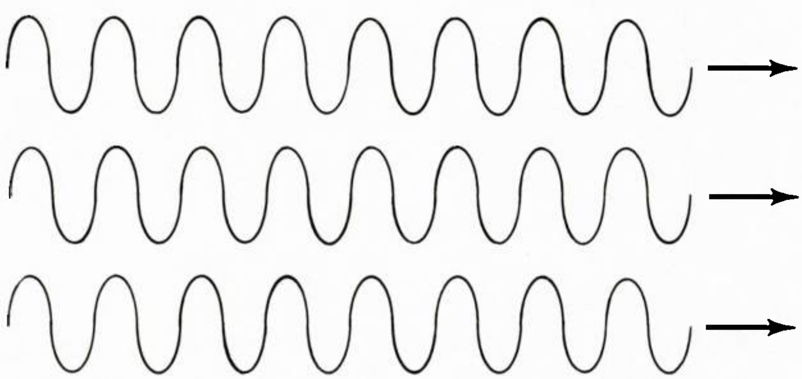


Figure 3. Monochromatic coherent radiation.

can be extremely large. Laser energy densities have been achieved which are billions of times greater than those found on the surface of the sun. These energy densities can vaporize any known substance on earth.

Laser radiations have been generated at nearly any desired wavelength throughout a considerable portion of the infrared spectrum, the entire visible light region, and well into the near ultraviolet region. Most lasers can be operated either continuously or in a pulse mode. Pulse lengths as short as a fraction of a picosecond (10^{-12}) have been demonstrated. Short pulses are desirable for application in such areas as ranging, optical information processing, high speed photography, optically generated plasmas, and many others.

TYPES OF LASERS

The information tabulated in Table 1 illustrates some selected properties of several typical lasers. The properties listed are representative values and not necessarily maximum or optimum.

TABLE 1
Select Properties of Some Typical Lasers

Type Laser	Class	Means of Excitation	Emission Wavelength (Angstroms)	Operation Mode	Output Power m – milliwatt M – megawatt	Efficiency %
Argon	Gas	Electric discharge	4579 to 5145	cw pulsed	0.01 to 10w (cw) 1 to 125 w (pulse)	0.01 to 0.05
Galium Arsenide	Semi-conductor	Injection current	9040	cw pulsed	1 w (cw) 100 w (pulse)	10 to 50
CO_2	Gas	Electric discharge	106,000	cw	0.1 to 10 Kw	5 to 30
Helium-Neon	Gas	Electric discharge	63280 11500 33900	cw	1 to 100 mw	.05 to 5
Krypton	Gas	Electric discharge	4619 4680 4762 4825 5208 5308 5682 6471 6764	cw	10 to 400 mw	.05 to 0.1
Ruby	Solid	Optical Pumping	6943	pulsed	1 to 350 Mw	0.01 to 2
YAG	Solid	Optical Pumping	10,600	cw pulsed	1 to 200 w	0.2 to 2

Solid State Lasers

The first laser, developed by Dr. Maiman in 1960, was a solid state device which used a synthetic pink ruby containing 0.05% chromium as an active medium. Relatively simple (see Figure 4), it consisted of a single crystal rod,

Figure 4. Solid state ruby laser.

approximately 1 cm in diameter and 10 cm in length. The ends of the rod were mutually parallel to within one minute of arc and silvered and polished to a mirror finish. One end was totally reflective, and the other end was approximately 90% reflective.

The ruby rod was surrounded by a helical zenon flashtube, a device used commonly in high speed photography. The purpose of the flashtube was to generate an intense flash of light to excite the chromium atoms to cause the population inversion. The flashtube was actuated by discharging a 100 micro-farad capacitor bank charged to approximately 2000 joules (watt seconds). The coherent radiations, which emerged through the partially reflecting end of the ruby rod, had a wavelength of 6943 Angstrom units (10^{-8} cm), which lies in the red frequency spectrum.

Although the ruby laser was the first solid state laser material to be developed, it remains one of the most important from the standpoint of power output and efficiency. Other solid state lasing materials include the alkaline earth tungstates and molybdates, fluorides, synthetic garnets, and optical glass lasers doped with rare earth ions.

Typically, the crystal or solid state lasers provide high energy outputs, and they require powerful light sources in order to excite the atoms to the point of radiation. Pulsed power outputs in excess of 10,000 megawatts have been demonstrated. However, the efficiency of present solid state lasers is low (usually considerably less than 1 percent), although efficiencies of several percent have been demonstrated. Furthermore, because of internal heating effects, high power solid state lasers generally are operated in a pulsed mode and are cryogenically cooled. Heat causes the bandwidth of the laser to broaden, and in some instances, causes the laser material to stop lasing and even shatter. Another reason for pulsed operation is the practical difficulty encountered in trying to provide a sufficiently powerful continuous source of light to activate and maintain the lasing action. Typical ruby laser pulse lengths range from a fraction of a nanosecond (10^{-9}) to a few milliseconds. Pulse lengths in the picosecond region have been achieved.

Once the synthetic ruby laser was developed, hundreds of other substances, including gases, liquids, plastics and semiconductor materials, have been made to lase. Whereas the ruby laser required an intense flash of light to initiate the lasing action, other lasers use electric current, chemical reactions, and even high temperature collision of gases to activate the lasing action.

Gas Lasers

Shortly after the first solid state laser was developed, lasing action was achieved using gas mixtures as the active laser medium. The configuration of an ordinary gas laser is also quite simple. It consists of a cylindrical gas discharge tube with reflecting mirrors at each end, forming an optical cavity resonator. Typically, an exciter or radio-frequency oscillator provides the energy to excite the gas or gas mixture and thereby activate the lasing action. Generally, gas lasers are operated at room temperatures.

Many gases have been found suitable for lasing: argon, carbon dioxide, helium, neon, nitrogen, krypton, zenon, and various mixtures of these gases.

When the coherent and monochromatic properties of the various types of lasers are compared, gas lasers exhibit the best performance. Their output radiation can be controlled to achieve narrow bandwidths and beamwidths. Spectral purities of better than one part in 10 billion in the narrowness of wavelength output have been achieved. Although the power output of gas lasers is low (with the exception of CO_2 and fluorine/hydrogen lasers), coherent *energy densities* millions and even billions of times greater than that found on the surface of the sun can be obtained. Such high intensities are, of course, limited to small areas. Another advantage of gas lasers is their ability to provide frequency outputs over a wide range of wavelengths. From Table 1, for example, the krypton laser produces laser radiations in the red, yellow, green and blue frequency spectrum. These output radiations can be produced either simultaneously or individually. Although the helium-neon laser is most often operated at a wavelength of 6328 Angstrom units, over 150 different laser transitions have been observed in the range from approximately 0.33 to 125 microns (10^{-6} meters).

Semiconductor Laser

The simplest and most efficient type laser is the semiconductor or diode laser. Although many semiconductor materials exhibit laser action, the most extensive research and development has been devoted to gallium arsenide (GaAs). The structure of the GaAs laser and most other semiconductor lasers is similar to that of an ordinary square or rectangular shaped p-n junction diode (see Figure 5). Small amounts of impurities are introduced into the semiconductor material by diffusion processes creating a positive and a negative (p-n) region. The junction between the positive and negative regions, normally approximately one micron thick, is the light emitting region.

Two sides of the junction region are cleaved or cut parallel and are highly polished, creating the optical resonant cavity. The remaining two sides are either roughly finished or offset angularly to avoid the generation of radiation in an undesired direction.

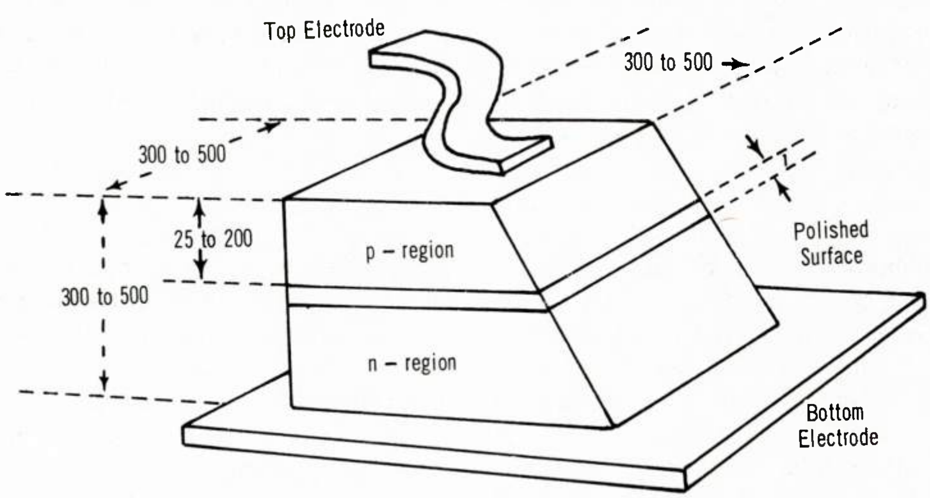


Figure 5. Schematic of a Ga As semiconductor laser (dimensions shown are in microns).

This type of laser can be energized by passing an electron current through the junction region. Electrons passing from the n-region drop into "holes" (atoms with a deficiency of electrons) in the p-region, emitting photons in the recombination process. If the current density is sufficiently high, stimulated emission occurs.

The overall efficiency of semiconductor lasers is relatively high. Efficiencies as high as 70% have been reported. On the other hand, the power handling capacity of these diode lasers is small because of the limited heat dissipation available in the junction. Although one watt cw outputs have been achieved at room temperature, cryogenic cooling becomes necessary to obtain reliable cw operation. This is accomplished by placing the diode laser in contact with a metallic heat sink immersed in a cryogenic coolant.

One major disadvantage of the semiconductor laser emissions is their large beamwidth, which can be as wide as ten degrees. However, the ease with which the output beam can be amplitude modulated (by varying the pumping or exciting electric current), combined with the overall high efficiency and small size of the laser, makes the semi-conductor laser an attractive device.

Liquid Laser

The fourth basic type is the liquid laser. For example, two classes of fluorescent organic compounds, the rare earth chelates and dye-molecules, form the basis for one class of liquid lasers. The active molecules of these compounds, each consisting of numerous atoms, are dissolved in ordinary liquid solvents such as water or alcohol. Most liquid lasers have been activated to lase either by the output beam of a ruby laser or by intense flashlamp excitation. Liquid lasers appear to have interesting possibilities, but their potential has not yet been realized because of the inefficiency of existing systems. One potential advantage of a liquid laser is the minimization of the heat dissipation problem. The heat added to the liquid by the pumping process and lasing action can be removed by circulating the liquid between the laser and a heat exchanger, instead of surrounding the entire device with an external and less efficient cooling system. Furthermore, like gas lasers, liquids allow complete freedom from the problems of growing and shaping large solid state crystals of laser quality. Suitable liquids can be produced in large quantities and placed into a container of any desired shape.

However, a more significant advantage of using the fluorescent organic compounds as liquid laser materials is that their output wavelengths can be changed by varying the concentration and/or composition of the active molecules. Thus, the liquid laser can be tuned over a large frequency range, providing virtually any wavelength throughout the visible frequency region as well as parts of the infrared and ultraviolet regions of the electromagnetic spectrum.

LASER APPLICATIONS

Although in this century the discovery and development of a totally new technology have not been uncommon, the laser appears to have been an especially significant event. From the time of its initial development, the intense, monochro-

matic and coherent properties of the laser created a fascinating interest among many scientists and engineers.

The one potential application that the word laser evokes in many minds is some kind of weapon or death ray. However, the potential use of the laser as a destructive or lethal device is but one of its possible uses. Its other potential uses stretch the limits of one's imagination. It touches almost every field of endeavor, including the fields of communication, chemistry, biology, astrophysics, and industrial processes.

The laser even can touch us in a very personal way. For perhaps, one's sight or the sight of a loved one might be restored or saved by a delicate laser operation, called a "retinal weld." The retinal weld already has become a routine operation in which the torn or injured retina is welded to its support by photo coagulation with a suitable short intense pulse of laser light. The laser also has shown some promise in the treatment of eye tumors and skin cancers. And, as a "light knife," the laser may someday make bloodless surgery possible.

Laser properties have made possible other fantastic accomplishments. For example, only two years after the laser was developed, a ruby laser, considerably smaller and less sophisticated than those now available, shot a series of pulses to the moon, 240,000 miles away. The laser beam illuminated a spot of less than two miles diameter and was reflected back to earth with enough strength to be measured by sensitive electronic equipment. As a comparison, the beam of the best achievable high quality spotlight, if it could reach that far, would spread to thousands of miles.

Spacecraft which have softlanded on the moon have photographed laser beams directed from earth to the moon using television cameras aboard the spacecraft. The pictures which showed small bright spots emanating from the dark side of the earth were then relayed back to earth. It is interesting to note that the laser outputs were less than three watts, illustrating how far a modest amount of power can go on a laser beam.

Obviously, it would be impossible to discuss all of the laser applications which have been achieved or which are under investigation. Therefore, the remainder of this section will be devoted to some of the major military and space applications.

The laser applications of particular interest to the military include the fields of communication, laser radar systems, surveillance, instrumentation, and weaponry.

Communications

One fruitful potential use of the laser is in the field of communications. The amount of information (information capacity) which can be encoded (modulated) on to an electromagnetic carrier is directly proportional to the frequency of the carrier wave. Because of the extremely high frequencies which lasers generate, enormous potential information capacities are feasible. For example, a one percent modulation bandwidth* operating on a 10^{15} Hertz laser carrier wave

* Modulation bandwidths of 10% are common at microwave frequencies.

provides a 10,000 gigahertz information bandwidth. This information bandwidth is several orders of magnitude wider than the total radio frequency spectrum.

Another significant advantage is the coherence properties of the laser which permits a highly collimated beam to be formed. This equates to high antenna gains and enables high data rate communications over vast distances. It appears within the state-of-the-art to obtain near real time television pictures at interplanetary distances of 100 million miles using laser techniques.

Because of the spectral purity of the laser radiations, effective optical filter techniques can provide a high degree of discrimination against naturally occurring background radiations from the sun, stars, and planetary bodies. The highly collimated laser beam also provides the capability of generating jam-resistant, secure communication links which are at times very desirable from a military viewpoint. It is possible to intercept a transmitted laser communications link only if the interceptor is located in a direct line with the transmitter and receiver.

Laser communications offer a possible solution to the radio blackout problem encountered by reentering spacecraft. Scientists already have demonstrated in laboratory experiments that laser radiations, unlike radio waves, will penetrate hot ionized gases or plasmas equivalent to reentry plasmas.

On the other hand, there exists several limitations in the use of the laser for communications. Laser radiations are subject to absorption and dispersion by atmospheric particles, particularly under conditions of rain, snow or fog. These atmospheric perturbing effects combined with the line of sight constraints limit terrestrial point-to-point communications to distances of some 30 or 40 miles. Because random outages cannot be tolerated in major communication systems, where even short interruptions in service cause chaos, only specialized links are candidates for terrestrial point-to-point circuits. Examples of such communication circuits would be secure or covert communication links and data links within nuclear test environments.

Although the line of sight problem can be removed by using repeaters to provide amplification and refocusing, the atmospheric effects cannot be eliminated. However, one means of circumventing the atmospheric aberrations is to beam the laser energy within hollow pipes which are either evacuated or filled with optical transparent and inert gases. These pipes, containing laser amplifiers (at intervals of approximately 50 to 100 miles) and reflecting surfaces to bend the laser beams where required, could enable the entire country to be spanned by laser communication networks similar to presently installed gas and oil pipelines. Several short distance experimental systems are already in existence in the United States as well as in France and the Soviet Union. Needless to say, pure space communications would not be affected by atmospheric aberrations (because of the near vacuum of space) or by line of sight limitations.

Other disadvantages of laser communications include the difficulties associated with pointing the very narrow transmitted laser beam and the problems of acquisition and tracking at thc receiving end. In addition, even disregarding the relatively low efficiency of most lasers, the efficiency of the receiving detection devices and other modulating and demodulating subsystems is considerably lower at laser frequencies than at radio and microwave frequencies.

Laser Radar

Numerous existing or potential laser device applications use the radar principle. These cover the gamut from simple hand carried range finders used for surveying or battlefield ranging to sophisticated space tracking and detection systems. Other potential laser radar functions include spacecraft signature analysis (target length, size and shape measurements), altimetry, automatic space rendezvous and docking, calibration of space track radars by simultaneous tracking of targets, and terrain clearance and obstacle avoidance.

In principle, laser radar does not differ from conventional microwave radar. Pulses of laser energy are directed toward a target, and the time elapsed between the pulse emissions and their reflected returns is measured. Range is obtained directly from the product of the known speed of the laser energy (the speed of light) and the time elapsed. The difference between the two systems is primarily a function of the higher laser frequency. The elements of the laser radar are merely the optical analogy of their microwave counterparts.

Although the range resolution of a pulsed radar system is primarily a function of the sharpness of the leading edge of the pulse, and the accuracy in measuring the two-way pulse transit time, the highly collimated laser beams enable range measurements in environments where conventional radars fail because of their wider beam widths. For example, in many cases conventional fire control radars in low flying aircraft become useless at angles of attack of less than approximately 15 degrees because of ground reflections. Laser radars enable angles of attack of only a few degrees. For the same reason, laser radars function far superior in many other ground point-to-point or near space operations. Thus, laser systems can provide tracking data during the launch and initial phase of flight of space vehicles where conventional radar is affected by ground clutter.

In a near earth space tracking role, radar can determine range to an accuracy of approximately 100 feet; the laser narrows the error to about 25 feet. Accuracies, using cooperative satellites equipped with special mirrored corner reflectors, are better than 10 feet at this distance. Successful ranging has been demonstrated on a cooperative satellite at 1000 miles range.

Laser radars are particularly effective in supplementing Baker-Nunn cameras, used in several satellite tracking networks, by providing range information in addition to illumination for photography when the spacecraft is in the earth's shadow. Since the Baker-Nunn camera provides angular coordinates, the additional range information specifies the position vector of the spacecraft at the time of observation.

Another advantage of the laser is its modest optical lens aperture requirements to concentrate its emitted beam. A one minute of arc beam can be obtained from an approximate one foot diameter lens. In comparison, the FPQ–6 radar, which is used in a space tracking role, requires a 30 foot diameter antenna to produce a 24-minute of arc beam. Angular resolution is a function of the narrowness of the radiated beam. Because of the narrow laser beam, present target azimuth and elevation angle resolution shows at least an order of magnitude of improvement over conventional radars.

However, as in laser communications, laser radar has several disadvantages with respect to conventional radar. These include the difficulty in acquiring the target and, for terrestrial applications, being exposed to the attenuations and aberrations caused by the atmosphere, particularly under adverse weather conditions. The difficulty in acquiring the target is especially severe in the large volume search problem, particularly with uncooperative targets.

Surveillance

The unique properties of lasers qualify its use for such surveillance functions as imaging, multicolor sensing, moving target indication, and covert photography illumination. Competing technologies are passive infrared, side-looking radar, and conventional photography. The advantages of the laser systems include better resolution than obtained by passive infrared and side-looking radar, and more covert night operation than conventional photography.

The covert nature of an active laser surveillance system is based on two factors: providing its own illumination and using a narrow pinpoint beam in a rapid sequential scanning process. Providing coherent self-illumination enables operation under day and night conditions. The highly collimated radiation permits a high degree of covert action even when using laser frequencies in the visible range. Use of frequencies in the invisible infrared or ultraviolet spectrum would further enhance covert operation.

The sharpness of the laser beam lends itself also to such special applications as terrain profile or contour mapping. Additionally, it is hoped that high resolution laser illumination will improve spacecraft signature analysis (target size, shape, attitude, and movement measurements).

Another important area in which the laser may improve present capabilities is in underwater surveillance. The inadequacy of the present systems is attested to by the difficulty the Navy has experienced in locating submarines and nuclear weapons lost in the sea. Although underwater attenuation of light transmission is well known to be several orders of magnitude above those of sound waves (sonar application), the use of laser energy in the blue-green region of the visible frequency spectrum provides hope that viable underwater surveillance systems can be developed. Sea water offers the least attenuation to these frequencies. In addition, the sensitivity of laser radiation detection devices is highest in this same frequency range. Present research is aimed at minimizing one of the greatest problems that needs to be overcome—back-scattering due to the suspended organisms and minerals, as well as the water itself.

Instrumentation

There are many interesting laser applications which fall under this heading, such as radar and gun bore sight alignment techniques, optical computers and data processing, wideband high resolution recording, techniques for measuring vibration of spacecraft undergoing simulated flight tests, and visual display technology which promises to produce cathode ray tube type screen sizes which are at least an order of magnitude greater than present displays.

Another fascinating area is holography, a means of storing (making a hologram) and reproducing, at will, a three dimensional image of an object or scene. Holograms can be produced in black and white or color. The reconstructed stored image is a virtual image, floating in space some distance behind the hologram. If the observer changes his viewing position, he can literally look around the object or elements in the scene. The broad range of potential applications of holography almost rivals its parent field of laser application.

Another potentially valuable application of the laser involves the laser gyro which does not depend upon the properties of inertia. The laser gyro is attractive because it provides a rugged and simple design with no moving parts—not even gimbles. It can be attached rigidly to a vehicle. Its other advantages include the ability to operate in extremely high acceleration and high turning rate environments, and the ability to perform almost instantaneously after being switched on.

The laser gyro is based on the principle that frequency differences will develop between two simultaneous contrarotating laser beams in a closed-loop optical cavity due to rotation of the optical cavity about its axis. This, in turn, is based on the fact that the frequency of operation of a laser is a function of the cavity length. If the optical path is stationary, the two contrarotating beams take the same time to traverse the loop, and their frequencies are the same. If the gyro is rotated about its axis, the effective path length for one beam is shortened while that of the other is lengthened, causing the frequency of one beam to increase and the other to decrease. The resulting frequency difference may be observed by optical heterodyning. The magnitude of the frequency difference is proportional to the rate of rotation. Three of these planar laser gyros can be positioned orthogonally to provide pitch, roll, and yaw information.

Weaponry

Ever since the laser was developed, there has been speculation that the laser would eventually develop into a death ray or weapon capable of destroying ballistic missiles at long range. There is no doubt that the intense collimated output energies of lasers are destructive and dangerous at short ranges. For example, blindness or severe eye damage can be produced using existing lasers at distances of several miles. The discussion here, however, will be limited to the laser's potential as a defensive radiation weapon against ballistic missiles.

As a defensive radiation weapon, the laser would have a significant advantage over incoming missiles. The velocity of laser energies is the same as that of light, 186,000 miles per second. Compare this with a velocity of ballistic missiles of approximately 17,000 miles per hour, or less than 5 miles per second. It becomes obvious that the laser energies would be traveling nearly 40,000 times faster than any incoming missile, and it follows that no lead angle for destroying the missile would be required.

Another advantage that the laser would have as a defensive radiation weapon is that it is a clean weapon. Present-day missile defense systems require nuclear warheads, and these warheads could cause radioactive fallout over friendly areas, particularly in a missile terminal defense situation.

The possibility of using lasers as defensive radiation weapons depends upon the amount of energy that can be transmitted to the target missile. In considering the amount of laser energy required to destroy an incoming missile, one must also consider that ballistic missiles are designed to accept enormous heat fluxes to survive reentry. Therefore, extremely high laser energies would be required to effect missile destruction. In addition, a considerable amount of the total laser energy available would be absorbed and dispersed by the atmosphere. In short, it would require laser energies far greater than any that have been produced under laboratory conditions.

On the other hand, the power requirements for missile defense lasers could be reduced significantly if the lasers were airborne or mounted on space platforms. One drawback of this approach, however, can be attributed to the fact that high power lasers are relatively inefficient devices and require extremely large power inputs. The capabilities for producing these power inputs for lasers under airborne or space conditions simply do not exist at this time, nor are they likely to exist in the immediate future.

Because of such factors, the development and use of the laser as a defense against ballistic missiles cannot be forecast with confidence at this time. It is obvious, however, that the military implications of such a system warrant extensive research and development.

Laser—the term does not connote science fiction. Some marvels of its application are in being today. We cannot yet imagine what its future has in store for us.

BIBLIOGRAPHY

"Applied Optics." *Proceedings of the IEEE,* Oct 1966.

BAKER, C. E. "Laser Display Technology." *IEEE Spectrum,* Dec 1968, pp. 39–50.

BIRNBAUM, G. *Optical Masers: Physical Principles.* New York: The Academic Press, 1964.

BROTHERTON, MANFRED. *Masers and Lasers: How They Work, What They Do.* New York: McGraw-Hill, 1964.

BURNS, F. P. "High-Power Lasers—Their Performance, Limitations, and Future." *IEEE Spectrum,* March 1967, pp. 115–120.

CARROLL, JOHN M. *The Story of the Laser.* New York: A. P. Dutton & Company, 1964.

CHANG, W.S.C. *Lasers and Applications.* A collection of papers from the Symposium on Lasers and Applications at Ohio State University, 1962. Columbus, Ohio: The Ohio State University Press, 1963.

DEMARIA, A.J. "Picosecond Laser Pulses." *IEEE Spectrum,* Jan 1969, pp. 2–25.

HERRIOTT, D.R. "Applications of Laser Light." *Scientific American,* Sep 1968, pp. 141–156.

KINSEL, J.S. "Lightwave of the Future: Optical PCM." *Electronics,* Sep 16, 1968, pp. 123–128.

KLEIN, H. ARTHUR. *Masers and Lasers: Construction and Use.* Philadelphia and New York: J. B. L ppincott Company, 1963.

KOLKER, E. B. "Deep Space Optical Communications." *IEEE Spectrum,* Jan 1967, pp. 75–82.

KORNSTEIN, E. "Blue-Green High Powered Light Extends Underwater Visibility." *Electronics,* Oct. 14, 1968, pp. 140–150.

LARACH, SIMON. *Photoelectronic Materials and Devices.* Princeton, N.J.: D. Van Nostrand Company, 1965.

LEMPICKI, A. "Liquid Lasers." *Scientific American,* June 1967, pp. 80–90.

LENGYEL, BELA A. *Lasers: A Generation of Light by Stimulated Emission.* New York and London: Wiley Press, 1962.

———. *Introduction to Laser Physics.* New York and London: Wiley Press, 1966.

———. "Evolution of Masers and Lasers." *American Journal of Physics,* Oct 1966, pp. 203–213.

LYTEL, ALLAN H. *ABC's of Lasers and Masers.* Indianapolis: H. W. Sams, 1963.
MELOY, T. "Laser's Bright Magic." *National Geographic,* Dec 1966, pp. 858–881.
MILLER, B. "Wideband Video in Digital Form Transmitted over Laser System." *Aviation Week,* Aug. 22, 1966, pp. 61 plus.
PATRUSKY, BEN. *The Laser.* New York: Dodd, Mead and Co., 1966.
PAY, R. "Laser Technique May Aid A-ICBM Systems." *Technology Week,* Nov 7, 1966, pp. 33–34.
PENNINGTON, K. S. "Advances in Holography." *Scientific American,* Feb 1968, pp. 40–48.
SALTONSTALL, ROBERT, Jr. *The Commercial Development and Application of Laser Technology: A Report.* New York: Hobbs, Dorman & Co., 1965.
SCHAWLOW, A. L. "Laser Light." *Scientific American,* Sep 1968, pp. 120–136.
STEHLING, KURT R. *Lasers and Their Applications.* Cleveland: World Publishing Co., 1966.
TOMIYASU, KIYO, ed. *The Laser Literature.* New York: Plenum Press, 1968.
VOLLMER, J. "Applied Lasers." *IEEE Spectrum,* June 1967, pp. 66–70.

CHAPTER 8

ATMOSPHERIC PENETRATION

NOT LONG AGO, no one knew whether it was possible to bring a vehicle safely back from space. Today successful atmospheric penetration is routine. The state-of-the-art in ballistic missile reentry vehicles has progressed rapidly and has reached the third generation of aerodynamic shapes. Lifting body craft, both manned and unmanned, have been developed and tested. Because the lifting body shape permits a high degree of maneuverability during reentry, its application to ballistic missile reentry vehicles and operational spacecraft will provide considerable flexibility in mission planning.

The rapid progress in the field of atmospheric penetration is based on a thorough knowledge of the factors involved. The purpose of this chapter is to enable the student to gain an appreciation for these factors, and it covers:

1. The characteristics of ballistic trajectories in the atmosphere.
2. The nature of aerodynamic and heating loads.
3. Some characteristics of lifting vehicles.

The intent is to limit the area of study to vehicles entering the earth's atmosphere from low earth orbit or to the reentry of ballistic missiles. In other words, emphasis is not on vehicles entering from deep space at very high velocities but rather on fundamentals, normally considering entry velocities of about 25,000 ft/sec.

In the vacuum of space, the trajectory of a vehicle is governed by its thrust, velocity vector, and gravitational force. It is modified in the atmosphere by aerodynamic forces which also cause heating due to air friction. Thus, as vehicles return from space flight, aerodynamic forces come into effect, and the relatively simple orbital relationships alone can no longer be used for predicting a trajectory. The equations of motion must consider the vehicle characteristics and nonlinear drag terms: atmospheric drag which is a function of velocity squared (v^2) and atmospheric density (ρ) which is a function of altitude.

BALLISTIC TRAJECTORIES

Assume that a manned vehicle has been in an elliptical orbit about the earth, and that thrust has been applied to change this orbit to intersect the earth's atmosphere. As the vehicle enters the atmosphere, aerodynamic drag affects the trajectory. This raises some important questions for those who must recover the space vehicle and

for the crew inside. The first question and the one considered in this section is this: what are the characteristics of ballistic trajectories in the atmosphere?

For an accurate analysis of a trajectory one must consider many factors, such as vehicle characteristics, atmospheric entry angle, atmospheric density, variation of density with altitude, wind, earth rotation, earth curvature, and gravity. Since some of the factors such as velocity and density are ever changing during the ballistic flight in the atmosphere, the analysis of a trajectory is considered in very small increments. Such computations are laborious when done manually; therefore, computers are used to analyze trajectories. Even the computer does not produce exact solutions because the predominant factors, velocity and density, are never known exactly. The purpose of this section is to analyze the forces that influence motion of a vehicle.

Geometry and Assumptions

The three forces acting on ballistic vehicles as they penetrate the earth's atmosphere are drag, weight, and centrifugal force. These forces and the geometry of the trajectory are shown in Figure 1.

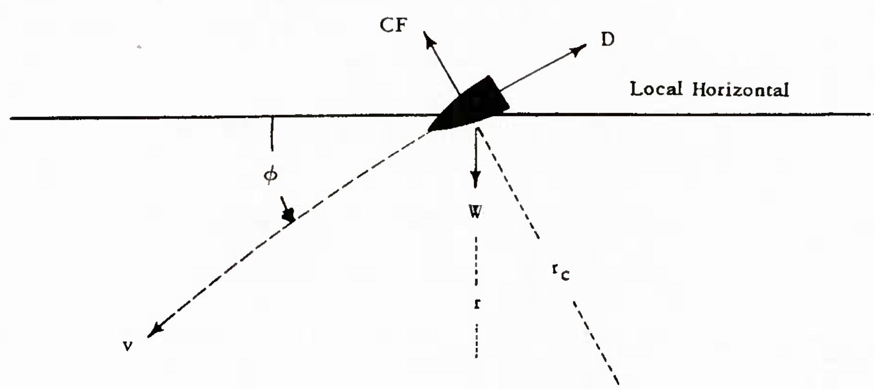


Figure 1. Geometry of a ballistic trajectory.

The vehicle is considered to be in ballistic flight through the earth's atmosphere when it is as shown in Figure 1. The vehicle has a velocity (v) at a distance (r) from the center of the earth. The direction of the vehicle as it enters the atmosphere is determined by measuring the angle which the trajectory makes with the plane of the local horizon. This is called the entry angle (ϕ). The radius of curvature (r_c) is perpendicular to the velocity vector at the point of consideration. The drag force (D) acts opposite to and along the velocity vector; and the weight of the vehicle (W) acts toward the center of the earth. Consequently, ballistic reentry is defined as entry in which the only effective external forces acting on the vehicle are aerodynamic drag and gravity.

Throughout the remainder of this chapter these assumptions will be made: (1) the earth is not rotating; (2) there is no wind; (3) the gravitational acceleration, g, is constant at 32.2 ft/sec^2.

Equations of Motion (Ballistic Trajectory)

To analyze the motion of a vehicle in flight through the earth's atmosphere, start with Newton's Second Law of Motion. This law states that the *unbalanced* force on a body is equal to the mass times the acceleration of the body:

$$F = Ma$$

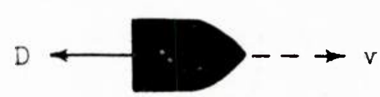
The force, *F*, is in lbs; *M* is in lbs-sec²/ft or slugs ($M = \frac{W}{g}$, where *W* is weight in lbs and *g* is the acceleration of gravity in ft/sec²), and *a* is the acceleration of the body in ft/sec². The change in velocity, Δv, divided by the corresponding increment of time, Δt, is equal to the acceleration:

$$a = \frac{\Delta v}{\Delta t}$$

FORCES ALONG THE DIRECTION OF MOTION.—During ballistic flight, the force, *F*, along the trajectory is an algebraic sum of 2 forces: the aerodynamic drag force (D) and a component of gravity ($W \sin \phi$).

1. The drag force acts opposite the direction of flight and can be computed from the basic aeronautical equation:

$$D = \frac{\rho v^2 C_D A}{2}$$



Where:

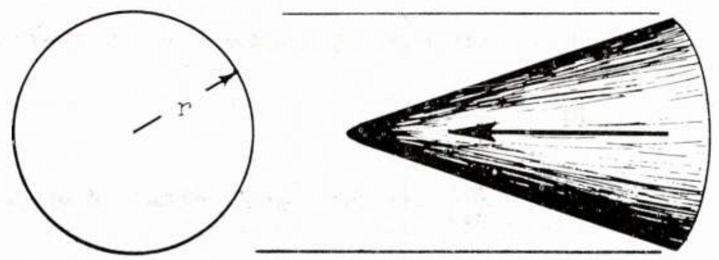
ρ = the density of the atmosphere and is expressed in slugs/ft³. At sea level ρ is 0.00237 slugs/ft³. (NOTE: The units of a slug are $\frac{\text{lbs-sec}^2}{\text{ft}}$ from $\frac{W}{g}$, so that $\rho = \frac{W}{(g)\ (ft^3)} = \frac{lbs - sec^2}{ft^4}$

v^2 = the velocity squared and the units are $\frac{ft^2}{sec^2}$.

C_D = the coefficient of drag and is a dimensionless number that reflects the shape of the vehicle. The more streamlined (less air resistant) vehicles have smaller C_D values than less streamlined vehicles. Actual values depend on many factors and vary with specific conditions. Typical values for the more streamlined shapes would be .5 and for the blunt shapes .8.

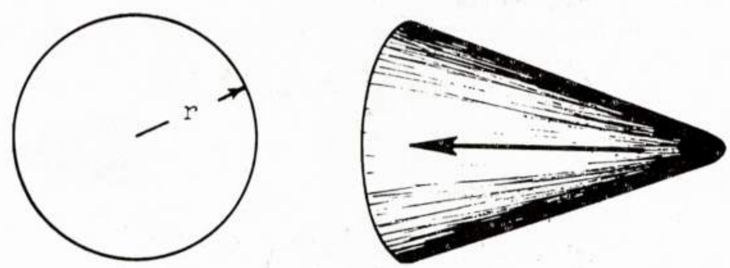
A = the effective frontal area of the vehicle exposed to the air stream and the units are ft^2. For example, the area of a sharp-nosed conical vehicle is the projected area:

$$A = \pi r^2$$

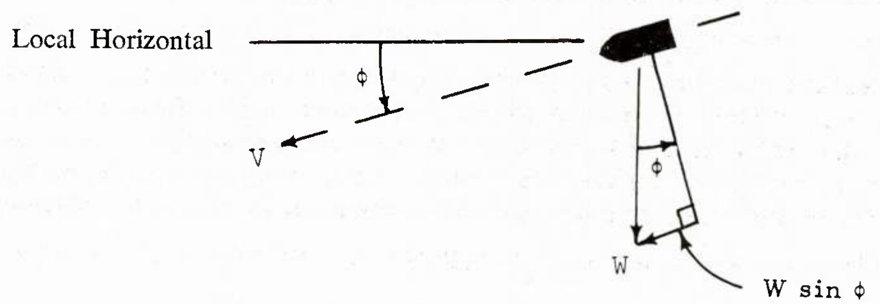


The effective frontal area of a blunt-nosed vehicle (conical) is also:

$$A = \pi r^2$$



2. The component of gravity *along the flight path* is in the direction of flight and, referring to Fig. 1, can be expressed as



W sin ϕ = component of gravity acting along the flight path, where W is the weight of the vehicle in pounds.

Now, substitute the above equivalents into Newton's Second Law, F = Ma, and develop an equation as follows:

$$F = Ma$$

$$F = -D + \text{gravity component} = -D + W \sin \phi$$

$$M = \frac{W}{g}$$

$$a = \frac{\Delta v}{\Delta t} \text{ (valid for a small segment of the trajectory)}$$

$$D = \frac{\rho v^2 C_D A}{2}$$

Now:

$$F = -\frac{\rho v^2 C_D A}{2} + W \sin \phi = \frac{W \Delta v}{g \Delta t}$$

Solving for $\frac{\Delta v}{\Delta t}$:

$$\frac{\Delta v}{\Delta t} = -\frac{g \rho v^2 C_D A}{2W} + \frac{g W \sin \phi}{W}$$

Simplifying:

$$\frac{\Delta v}{\Delta t} = g\left[-\frac{\rho v^2}{2\left(\frac{W}{C_D A}\right)} + \sin \phi\right] \tag{1}$$

The above equation determines the change in velocity along a small segment of trajectory. It indicates that for a small time increment, Δt, there is a corresponding change in the velocity of the vehicle. The magnitude of the Δv is a function of all the other values shown. Some of these other values (ρ, sin ϕ, and v) are ever changing during the flight.

A very important parameter in Equation 1, and appearing for the first time, is the *ballistic coefficient,* $\frac{W}{C_D A}$. This is a common parameter that appears in all studies of atmospheric penetration and plays a very important role in the behavior of a vehicle. It is simply the ratio of vehicle weight to a factor describing the degree of streamlining (C_D) and the effective frontal area of the vehicle as defined above. For example, a streamlined ballistic reentry vehicle (R/V) has a very large ballistic coefficient (1,000 lbs per square foot or more) when compared to that of the blunt-nosed Apollo capsule (which has a $\frac{W}{C_D A}$ of approximately 100 pounds per square foot).

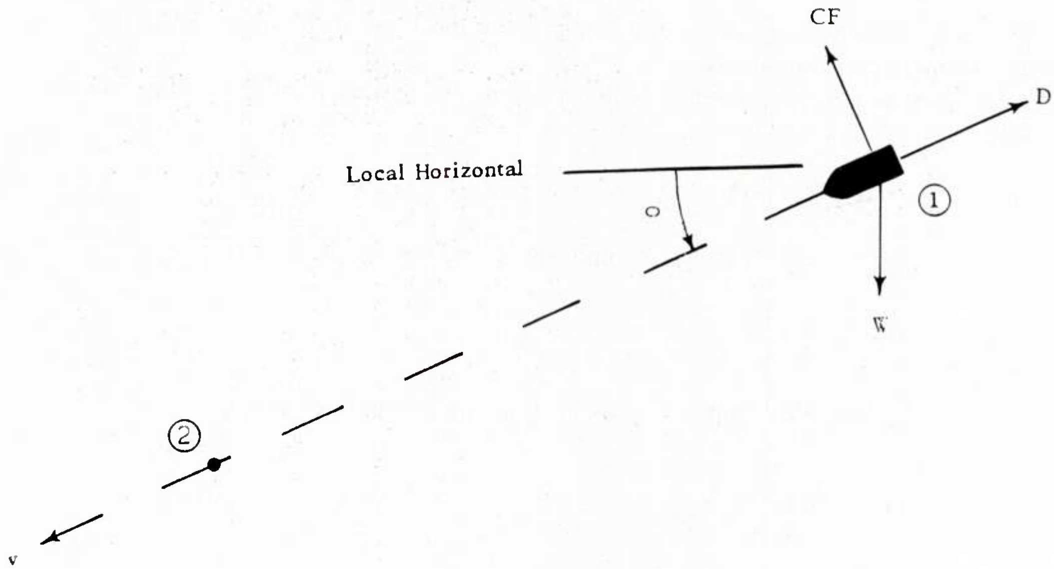


Figure 2. A segment of a trajectory.

To appreciate the complexity of the computation, examine a small segment of a trajectory. Assume that the value of the change in the velocity between point 1 and point 2 in a trajectory is desired.

In the small segment (point 1 to point 2) shown in Figure 2, ρ and ϕ do not change significantly, and the trajectory is assumed to be a straight line. Thus, Equation 1 can be written,

$$\Delta v = \Delta t \, g \left[- \frac{\rho_1 v_1^2}{2\left(\frac{W}{C_D A}\right)} + \sin \phi_1 \right] \qquad (2)$$

Example:

Equation 2 can be used to analyze the velocity change when the conditions are known at point 1. An example should clarify the equation and its use. Consider a manned vehicle entering the earth's atmosphere under the following conditions:

Given:

$$\Delta t = 1 \text{ second}$$

$$g = 32.2 \text{ ft/sec}^2$$

$$\text{altitude} = 480{,}000 \text{ ft (80 NM)}$$

$$\rho_1 = (4.12)(10^{-12}) \text{ slugs/ft}^3 *$$

$$v_1 = 23{,}000 \text{ ft/sec}$$

$$\frac{W}{C_D A} = 100 \text{ lbs/ft}^2$$

$$\phi_1 = 5 \text{ degrees } (\sin \phi = .087)$$

* ARDC Model Atmosphere.

Solve for: Δv, where $\Delta v = v_2 - v_1$

Substituting into Equation 2,

$$\Delta v = (1)(32.2)\left[-\frac{(4.12)(10^{-12})\ (23{,}000)^2}{(2)\ (100)} + .087\right]$$
$$= 32.2\ [-.0000109 + .087] \approx 32.2\ [.087]$$
$$= +\ 2.7 \text{ ft/sec}$$
$$v_2 = v_1 + \Delta v = 23{,}000 + 2.7 = 23002.7 \text{ ft/sec}$$

These calculations demonstrate that for this particular vehicle ($\frac{W}{C_DA} = 100$) and the low atmospheric density, the magnitude of the drag force is insignificant at high altitudes (80 NM). In fact, the drag force is so low that the velocity is increasing by a very small amount due to the small gravity component.

HEATING AND DECELERATION

Heating rates and deceleration loads are the most important effects of entry into the atmosphere. These effects are most severe when there is a combination of high atmospheric density and high vehicular velocity. The most critical condition for a vehicle upon entry would involve steep entry angle and high approach velocity.

A vehicle approaching the earth's atmosphere possesses a great amount of potential energy (PE) and kinetic energy (KE). While penetrating the earth's atmosphere, the vehicle loses very little PE in relation to the loss in KE; therefore, all the energy which is lost will be considered a KE loss. KE is a function of v^2. Thus, due to the high entry velocities, large amounts of KE must be dissipated as the velocity is reduced to a safe impact value for manned vehicles. It is vital that the energy be dissipated in order to prevent destruction of the vehicle or its payload.

The velocity of the vehicle at atmospheric entry depends on the mission. If entry is initiated from near earth orbit, then the velocity at atmospheric entry closely approximates that of the orbital velocity. The velocity of a reentering ICBM warhead is approximately the same as the velocity at burn-out of the booster. Entry from lunar missions involves velocities of 35,000 to 36,000 ft/sec. Regardless of the mission, the entry velocity will be very high and can be changed a significant amount only by *very large retro-thrust forces*. This would lead to severe payload penalities. On the other hand, small retro-thrust forces which allow a shallow entry angle can be used at a lesser cost in terms of weight. The atmosphere, instead of large retro-thrust forces, is used to slow the vehicle. A shallow entry angle tends to limit to high altitudes the region of high velocity. However, there are other factors influenced by the entry angle. For example:

1. Atmospheric peak heating rates and vehicle peak deceleration loads are less with shallow entry angles.
2. Vehicle heat loads may be greater for shallow entry angles.
3. Impact accuracy is less with shallow entry angles.

The heating and deceleration loads are a function of entry angle, velocity, and density. The remainder of this section will deal with the interrelationship among these factors and their influence upon heating and deceleration.

Heating

Assume a manned vehicle enters the earth's atmosphere at 24,000 ft/sec and uses atmospheric braking. Theory and flight have shown that the temperature that will be encountered would destroy a vehicle constructed of present-day metals. The minimum vehicle surface temperature that can be expected during peak heating, even under optimized configuration and entry angle, will be about 3,000°F. The temperature could reach 20,000°F. for steep entry angles. Two qucstions now arise: what heats the vehicle, and how is the vehicle protected from this intense heat?

During atmospheric penetration, KE is transformed by aerodynamic drag into thermal energy, heating the air surrounding the vehicle. Simply, the penetrating vehicle gets hot because it is close to the heat source—the air. The amount of heat transferred to the vehicle depends on the characteristics of the air flow near it. If all the heat in the air were transferred into the vehicle, it would be more than enough to vaporize it unless the vehicle were constructed of a material such as carbon.

A rigorous approach to the subject is too complicated and theoretical for the purposes of this text. In the following paragraphs a less complex method which will give sufficient accuracy will be used.

ATMOSPHERIC HEATING RATE.—Theory is that the heating rate is a function of the change in KE of the vehicle. The rate of KE change can be evaluated by using the equation for KE, which is:

$$KE = \frac{M\ v^2}{2}$$

In a small segment of the trajectory, the change in KE (ΔKE) per unit of time is:

$$\frac{\Delta KE}{\Delta t} = \frac{KE_2 - KE_1}{\Delta t}$$

$$KE_1 = \frac{Mv_1^2}{2}$$

$$KE_2 = \frac{Mv_2^2}{2} = \frac{M\ (v_1 + \Delta v)^2}{2} = \frac{M\ (v_1^2 + 2v_1\ \Delta v + \Delta v^2)}{2}$$

$$\text{SINCE } \Delta v^2 << v_1$$

$$KE_2 = \frac{M\ (v_1^2 + 2v_1\ \Delta v)}{2}$$

$$\therefore \frac{\Delta KE}{\Delta t} = \frac{Mv_1^2 + 2Mv_1\ \Delta v - Mv_1^2}{2\Delta t} = \frac{Mv_1\ \Delta v}{\Delta t}$$

It is now convenient to relate the above equation to an expression containing the aerodynamic drag term by use of Equation 1:

$$\frac{\Delta v}{\Delta t} = g \left[- \frac{\rho\ v^2}{2} \left(\frac{C_D A}{W} \right) + \sin\ \phi \right]$$

In the area of high energy conversion, where Δv is high, $\sin \phi$ is very small in relation to other parameters and can be neglected; this simplifies the above equation to:

$$\frac{\Delta v}{\Delta t} = - \frac{\rho v^2}{2} g \left(\frac{C_D A}{W}\right)$$

Multiply both sides of the above equation by M v and obtain:

$$\frac{M v \Delta v}{\Delta t} = - \frac{\rho v^3}{2} g \left(\frac{C_D A}{W}\right) M = - \frac{\rho v^3}{2} (C_D A)$$

Now equating the two terms which equal $\frac{M\, v\, \Delta v}{\Delta t}$:

$$\frac{\Delta KE}{\Delta t} = - \frac{\rho v^3}{2} (C_D A) \tag{3}$$

Surface heating rate. Equation 3 is an expression that shows thc *heating of the air,* as transformed from KE, at a *point* in the trajectory. Thus, for a given configuration (a known $C_D A$), and known velocity and altitude (density) history during a penetration, the equation can be used to determine the heating *trend* and the *point (altitude and velocity) of maximum heating.* A plot of these values (vs time or altitude) will depict the heating trend and reveal the area of peak heating.

No attempt is made in this text to compute the actual heating rates or temperatures. Such computations involve a highly technical area with limited "state-of-the-art" techniques and experience to date. There are many theoretical studies to which the student may refer. This text is concerned with the predominant heating factors.

Equation 3 shows the thermal energy transferred to the air; interest lies in knowing how much of this heat is transferred to the vehicle. It is important that the air, not the vehicle, absorb as much of this heat as practicable. The *energy conversion factor* (ECF) is the fraction of converted KE which enters the vehicle as heat. The heat input determines the type of vehicle protection requircd and the corresponding payload weight penalty imposed. This fraction of converted energy is as follows:

$$\frac{\text{Heat absorbed by the vehicle}}{\text{Total heat generated}} = \text{ECF}$$

The magnitude of the ECF depends primarily on the (1) vehicle shape, (2) velocity, and (3) atmospheric density (altitude). The effect of shape and density on the ECF and the type of air flow for typical ballistic trajectories is shown in Figure 3. This figure indicates that as much as one-half of the total heat generated would be transferred to the vehicle at very high altitudes by molecules of air impinging directly on the surface and transferring heat energy to the vehicle. Fortunately, at these extremely high altitudes, the atmospheric density is very low; thus, the KE transfer is correspondingly low. Large ECF's may also occur at lower altitudes for slender shapes.

The details shown in Figure 4 reveal why the shape of the vehicle has such strong influence on the ECF. This figurc shows the air flow pattern around a blunt (high drag) vehicle and around a streamlined (low drag) vehicle.

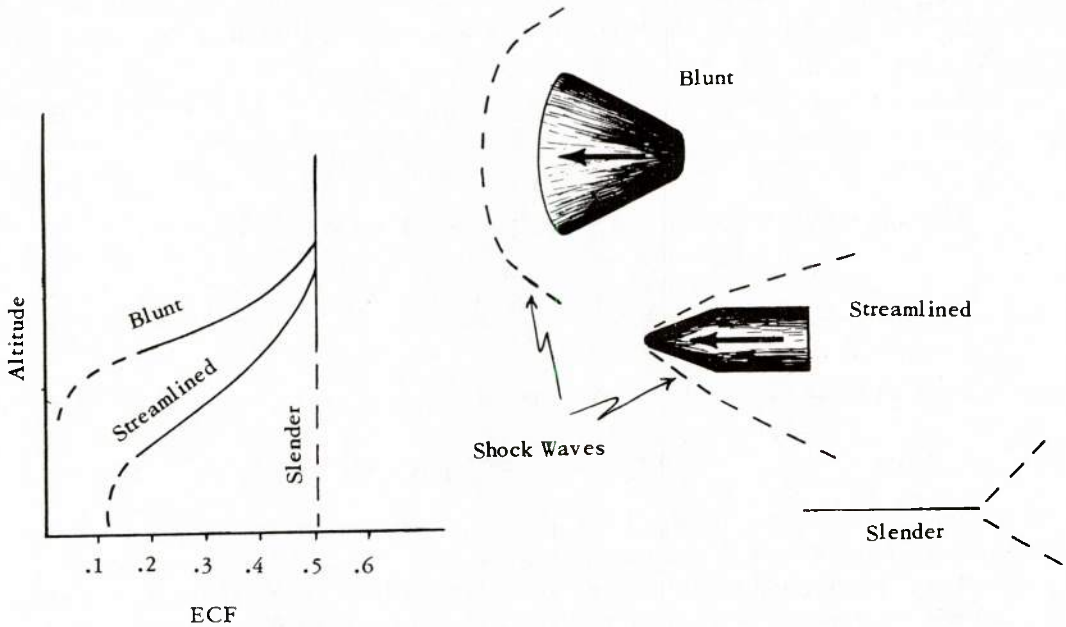


Figure 3. Parameters affecting energy conversion factor.

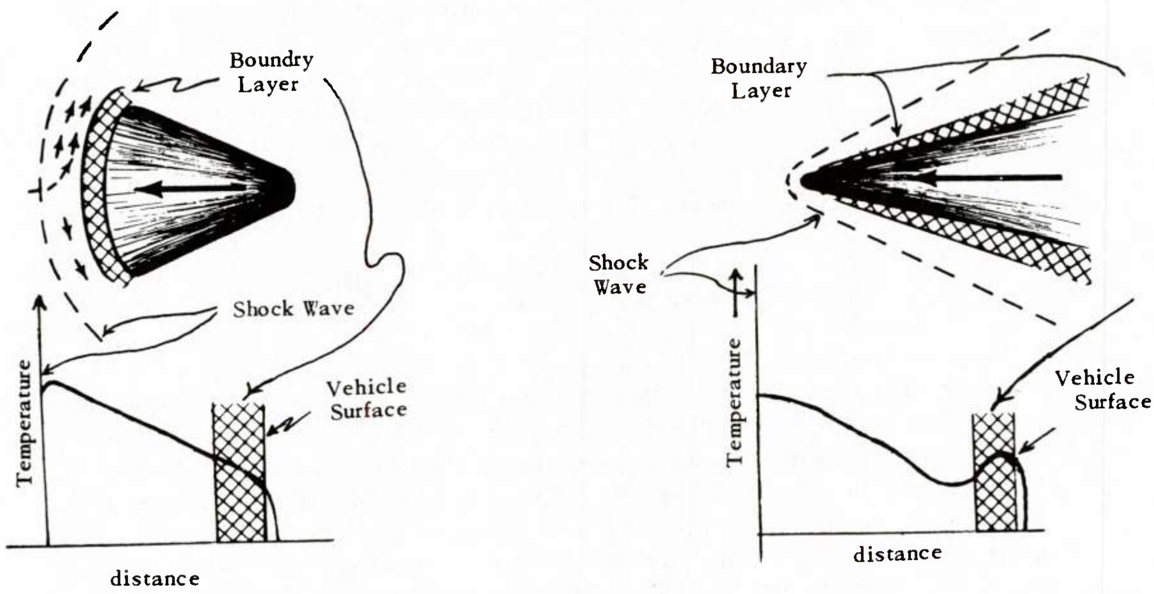


Figure 4. Aerodynamic heating.

With the *blunt* vehicle, the shock wave is detached and ahead of the vehicle. It is nearly normal (perpendicular) to the velocity vector. The air between the surface and the shock wave is hot and moves slowly with respect to the vehicle. The velocity of the air in the boundary layer (a thin layer of air next to the vehicle) is very low. *The boundary layer acts as insulation;* however, some heat is transferred from the hot air through the boundary layer to the surface of the vehicle.

With a *streamlined* vehicle, the shock wave is almost parallel to the air stream, and the air in the boundary layer is moving relatively fast. There is a large change in velocity (velocity shear) between the vehicle surface and outer edge of the boundary. *The viscosity of the air and the shearing action* cause the maximum temperature to occur within the boundary layer.

The rate of heat transfer is different for vehicles of different configuration. It can be seen that the temperatures created by the streamlined vehicle are high near the surface; thus, more heat is transferred to the streamlined shape than would be in the case of the blunt vehicle. This is the reason the ECF is higher for the slender vehicle.

The ECF is also influenced by the character of the air flow in the boundary layer—whether laminar or turbulent. In laminar flow, the air moves smoothly in layers (lamina). Turbulent flow has the characteristic of irregular, eddying, or fluctuating flow. The turbulent boundary layer occurs in the lower atmosphere and allows a much higher rate of heat transfer to the vehicle.

A typical variation of the ECF for a blunt vehicle penetrating the earth's atmosphere is shown in Figure 5. The ECF varies from about 0.5 at high altitudes to

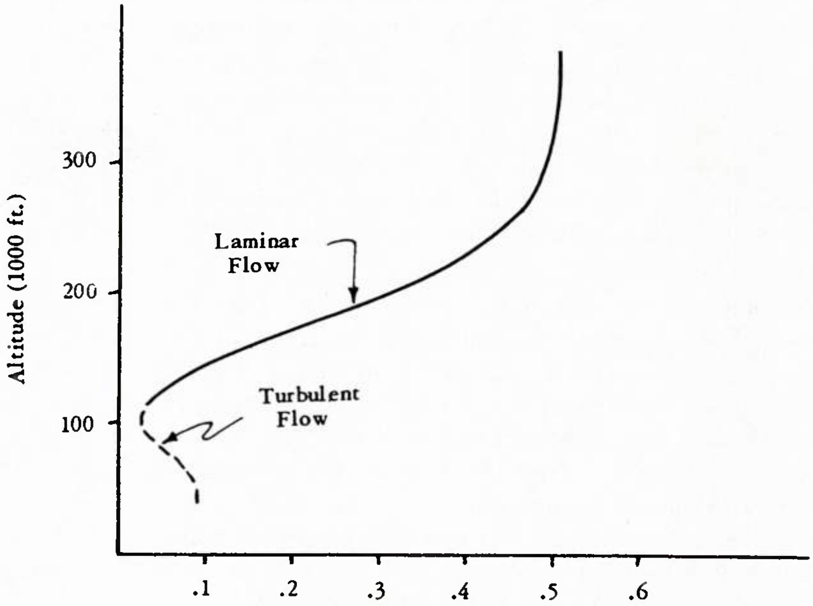


Figure 5. ECF vs altitude.

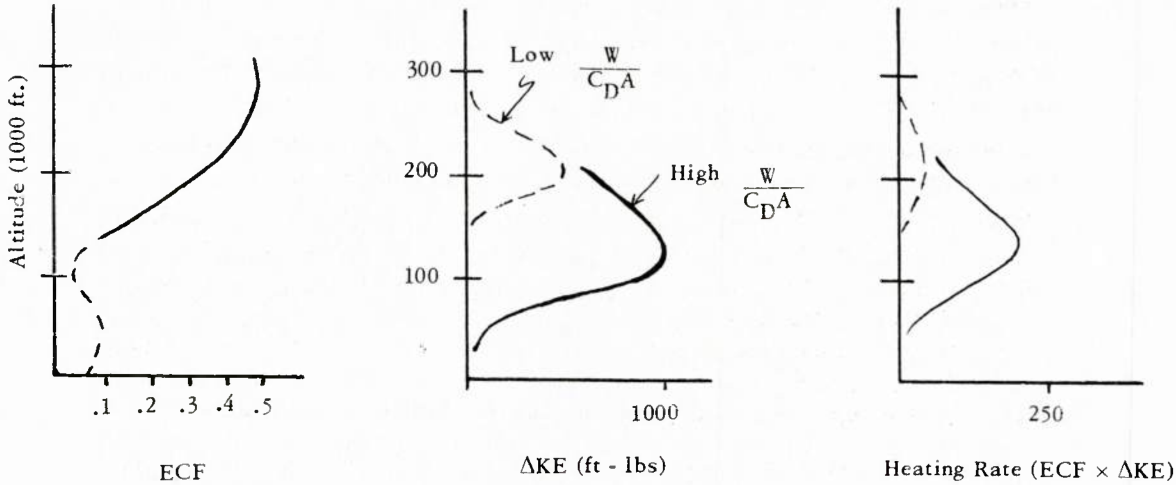


Figure 6. Influence of changes in KE and ECF on surface heating rate.

about 0.01 at 100,000 ft. At lower altitudes, the ECF increases again due to a turbulent boundary layer condition, as indicated by the dashed portion of the curve.

The rate at which heat enters the vehicle is a product of the rate of change in KE and the ECF. The influence of these factors is shown in Figure 6. The rate of change in KE is more dominating than the variance of the ECF. For example, at high altitude where the ECF is high, the rate of change in KE is low and the surface heating rate is low. At medium altitude (approximately 100,000 ft) where the ECF is low, the rate of change in KE is high and the surface heating rate is high.

A light-weight vehicle (low ballistic coefficient), or a lifting vehicle, will decelerate in the upper atmosphere and will have low surface heating rates. This will occur in spite of the high ECF because low density causes a small change in the KE. A heavy, streamlined vehicle (high ballistic coefficient) decelerates in the lower atmosphere and has a large surface heating rate. This will occur in spite of the lower ECF because the higher atmospheric density at the lower altitudes causes a high rate of conversion of KE.

TOTAL HEAT LOAD.—The term *"total heat load" means all the heat absorbed* by the vehicle during entry. An important point here is that the total heat input depends not only on the heating rate but also on the *duration* of the heating. A vehicle exposed to a low heating rate over a long period of time *may* absorb a larger total heat load than a vehicle exposed to a high heating rate for a short time. The vehicle designer's choice will lie somewhere between two extremes: a low total heat load caused by high surface heating rates for a short time, and a high total heat load caused by low surface heating rates over a long period of time. As one might expect by reference to Figure 6, *a lower total heat load may be obtained by programming the maximum change in KE at the minimum level of ECF.*

PROTECTION FROM HEAT.—A great many techniques have been proposed for protecting the vehicle from the intense heat during the atmospheric penetration. All these techniques fall into one of two general classes: radiation cooling and heat absorption systems.

If the surface heating rate is low, some of the heat may be radiated back into the atmosphere—radiation cooling. This type of cooling is limited by the maximum temperature that the vehicle is able to withstand. A point is reached where the amount of heat absorbed by the vehicle equals the amount radiated back into the atmosphere. For a given reentry condition and altitude an equilibrium temperature then exists, and the temperature of the structure has reached a maximum. Radiation cooling was the method used for protecting the lifting surfaces of the Asset glider. Such vehicles decelerate slowly and have high velocities for a long period of time at high altitude; consequently they undergo low heat rates but high heat loads.

A heat sink system uses a large mass of metal (usually copper) to absorb a great quantity of heat before melting. This technique was used early in the ballistic missile reentry vehicle program, but it proved too heavy for today's operational vehicles.

Transpiration implies the boiling off of a liquid. Water, for example, boils at about 70°F at the low pressures encountered in space. Therefore, it would maintain the part of the vehicle which it was protecting at 70°F. However, as the pressure increased during reentry, the boiling temperature would also increase.

Ablation is the wearing away, melting, charring, or vaporizing of the surface material. The vaporizing process leaves a thin vapor layer near the vehicle, thereby insulating the vehicle from the intense heat. Also, some heat is carried away from the vehicle by the wearing away and melting process. Examples of ablative materials are reinforced plastics, resin, fiberglass, and cork.

Deceleration

A spacecraft returning from a low earth orbit enters the earth's atmosphere at 24,000 ft/sec. The velocity of the vehicle must be greatly reduced prior to landing. The reduction of velocity per unit of time is the rate of velocity change, or acceleration. If the velocity is decreasing, the acceleration has a negative value and is called deceleration, the case when a ballistic vehicle penetrates the atmosphere. The larger the change in velocity per unit time, the greater the deceleration. Therefore, consider the very high change in velocity and corresponding deceleration loads that may range up to hundreds of "g's" with the peak "g" determined by the entry conditions. Knowing that each system has a maximum allowable "g," the following question arises: what factors influence the peak deceleration loads during penetration?

The motion of a vehicle during atmospheric penetration depends on (1) the vehicle's velocity, (2) atmospheric entry angle, (3) ballistic coefficient, (4) lift characteristics, and (5) the atmospheric density. During atmospheric entry the vehicle may pitch, yaw, roll, and experience accelerations in the longitudinal and lateral directions. Because it is of greater magnitude, the longitudinal deceleration of the vehicle is the only dynamic motion considered in this text.

The most important question of vehicle dynamics is this: what is the magnitude of maximum deceleration compared with the earth's gravitational acceleration, g? To determine the answer, again use Equation 1 to plot the deceleration history.

$$\frac{\Delta v}{\Delta t} = g\left[-\frac{\rho v^2}{2\left(\frac{W}{C_D A}\right)} + \sin\phi\right]$$

For lifting vehicles, the value of C_D can be obtained from a curve of C_D vs C_L. The maximum deccleration for a lifting vehicle depends on the entry conditions and the lift-drag ratio (L/D). The velocity decrease is extended over a longer period of time than for a ballistic vehicle, with consequently lower deceleration. Thus, both a shallow entry angle and aerodynamic lift serve to decrease thc deceleration during atmospheric penetration. Figure 7 shows the maximum longitudinal deceleration as a function of initial entry angle and L/D. Note that a large reduction in maximum deceleration can be obtained by relatively small L/D values. The longitudinal deceleration of a lifting vehicle is very low and is not a problem; therefore, the remainder of this section is concerned with ballistic (non-lifting) vehicles.

As stated in the previous scction on *Heating,* the magnitude of hcating (specific values of temperature and BTUs) cannot be accurately computed; however, the magnitude of deceleration can be easily computed. Deceleration is $\frac{\Delta v}{\Delta t}$ in Equation 1, and the units are in ft/sec^2. Dividing $\frac{\Delta v}{\Delta t}$ by "g" converts the magnitude of deceleration into terms of gravitational acceleration. Since the basis has been laid for computing deceleration, now return to the examplc problem and compute the deceleration for the manned ballistic vehicle entcring the earth's atmosphere.

$\Delta t = .1$ second was chosen, and $\Delta v = 2.7$ ft/ sec was computed. $\frac{\Delta v}{\Delta t} = \frac{2.7}{1} = 2.7$ ft/sec^2 (positive value; thus, vehicle is accelerating).

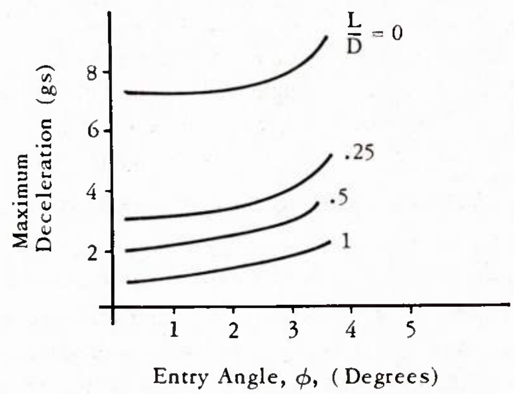


Figure 7. Maximum deceleration vs. entry angle, for various values of $\frac{L}{D}$.

In terms of the gravitational acceleration:

$$\frac{\Delta v}{g\,\Delta t} = \frac{2.7}{(32.2)\,(1)} = .08g$$

Consider now the same manned vehicle at 100,000 ft altitude travelling 8,000 ft/sec at an entry angle of 10 degrees under the following conditions:

$$g = 32.2 \text{ ft/sec}^2$$

$$\rho = (3.2)\ (10^{-5}) \text{ slugs/ft}^3$$

$$v = 8{,}000 \text{ ft/sec}$$

$$\frac{W}{C_D A} = 100 \text{ lbs/ft}^2$$

$$\phi = 10 \text{ degrees } (\sin\phi = .174)$$

$$\frac{\Delta v}{\Delta t} = g\left[-\frac{\rho v^2}{2\left(\frac{W}{C_D A}\right)} + \sin\phi\right]$$

$$= g\left[-\frac{(3.2)(10^{-5})\ (8000^2)}{(2)\ (100)} + .174\right]$$

$$\frac{\Delta v}{\Delta t} = g\ (-10.2 + .174)$$

$$\frac{\Delta v}{\Delta t} = -10.026\ g \text{ (negative value } \therefore \text{ vehicle is decelerating)}$$

EFFECT OF ENTRY ANGLE ON DECELERATION.—By analyzing the above, it is evident that the direct contribution of sin ϕ is negligible in the area of high deceleration; even at a 90-degree entry angle the maximum acceleration due to gravity would be 1 g (sin ϕ = 1). However, the entry angle shows its real influence indirectly. An examination of Equation 1 indicates that the entry angle is the only variable in the equation which can be controlled realistically. The other factors (gravity, velocity magnitude, atmospheric density, and the ballistic coefficient) are all fixed, and only the entry angle can be changed for a specific mission.

Theoretically, for any *fixed* entry angle and velocity magnitude, the maximum deceleration is also fixed. The only thing that can be changed is the altitude at which that maximum occurs, and that altitude can be changed only by changing the ballistic coefficient, $\frac{W}{C_D A}$. Vehicles with large ballistic coefficients merely penetrate further into the atmosphere before experiencing maximum deceleration. In other words, two different vehicles (regardless of difference in shape, size, and weight), approaching the atmosphere at the same velocity magnitude and angle, will experience the same maximum deceleration—only their trajectories and altitude of maximum deceleration will be different. Figure 8 shows the influence of changing the ballistic coefficient for a shallow entry angle (in the order of 1 degree). Note that both vehicles experience the same magnitude of deceleration; the basic difference is that the vehicle with the smaller ballistic coefficient encounters maximum deceleration at a higher altitude.

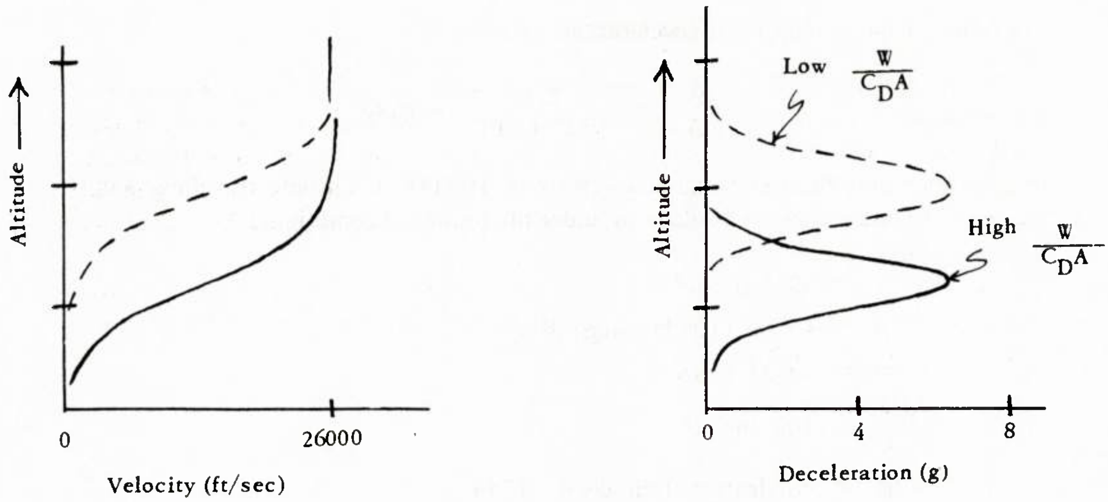


Figure 8. Influence of ballistic coefficient for given entry angle.

Theoretically, a ballistic reentry vehicle entering the earth's atmosphere with a given velocity magnitude has a peak deceleration which is determined only by the entry angle. If the entry angle is increased, the magnitude of maximum deceleration also increases. In other words, the steeper the angle, the higher the maximum deceleration. Figure 9 shows the effect of entry angle on the maximum deceleration for a vehicle such as the Project Mercury capsule. The shallow entry angles cause the deceleration to start higher in the atmosphere and last over a long period of time. This results in a smaller peak deceleration. If a vehicle with a smaller ballistic coefficient were used, these curves would shift to a higher altitude with identical peak decelerations for the corresponding entry angles. Conversely, a vehicle with a larger ballistic coefficient would cause the curves to shift downward.

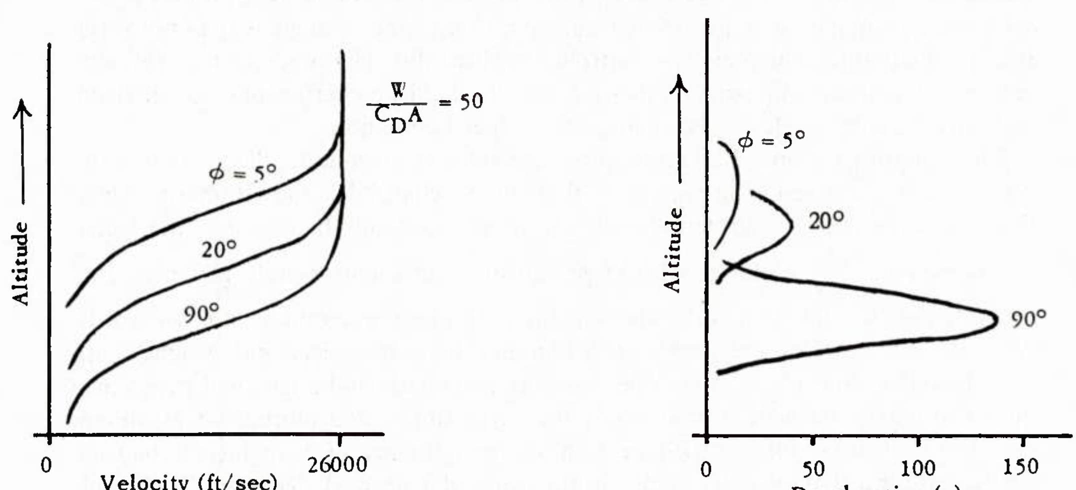


Figure 9. Influence of entry angle on velocity and deceleration.

LIFTING VEHICLES

The discussion thus far has been centered mainly on ballistic vehicles. These vehicles have an inherent disadvantage, especially when viewed from a military standpoint: the inability to control the vehicle through the atmosphere to arrive at an exact impact point. For example, after retrothrust has been used to deorbit a vehicle, there is little that can be done to control the final destination point of the vehicle—its impact point depends solely on ballistic parameters.

The lifting vehicle allows a more gradual descent through the atmosphere. This lifting capability leads to several potential advantages depending on the specific mission of the vehicles, such as:

1. Low deceleration loads.
2. Low heating rates.
3. Ability to maneuver in the atmosphere.
   a. Ability to land at a predetermined point.
   b. A landing "footprint."
   c. Minimum recovery forces required.
   d. A deep atmospheric entry corridor.
   e. Capability of synergetic plane change.

There are also some disadvantages:

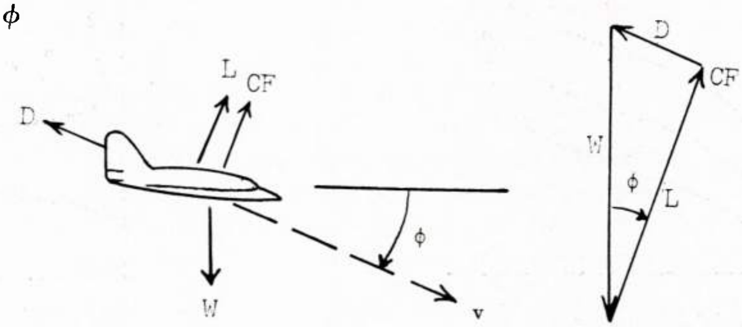
1. Lifting surfaces add weight which reduces the overall range and/or useful payload for a given booster.
2. The total heat load of the vehicle may be much greater due to the long period of deceleration.
3. Lifting technology lags behind ballistic technology (there has been far more experience with ballistic than lifting vehicles).

The above lists reveal some of the general characteristics of a lifting vehicle, not all of which will be discussed here. Rather, the purpose of this section is to explore the ground range capability of a lifting vehicle after it enters the earth's atmosphere.

In analyzing the glide of a lifting vehicle, consider the four forces acting on the vehicle during penetration: (1) weight, (2) drag, (3) lift, and (4) centrifugal force. Since the motion is unaccelerated by thrust, the equilibrium conditions lead to the following equations:

$$L + CF = W\cos\phi$$

$$D = W\sin\phi$$



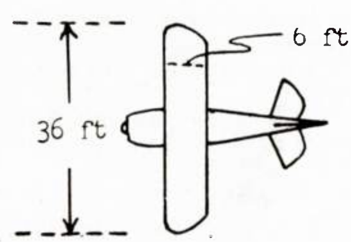
Previous discussion has defined all of these terms except the lift (L), which is the prime consideration in the lifting vehicle. A basic aeronautical equation is used to determine lift:

$$L = \frac{\rho v^2 C_L S}{2}$$

All the above parameters have been previously discussed cxcept the lift coefficient, C_L, and the area, S. The lift coefficient is a dimensionless number and reflects the lifting capability of a particular surface at a given angle of attack relative to the air-stream. C_L is also a function of the shape of the lifting surface. For example, a very high C_L is associated with thick wings which have high lift for a given wing area. These are usually found on slow-flying aircraft. High-speed aircraft cannot use thick wings due to the excessive drag caused by the thickness. Thin wings characteristically have a very low C_L, and they normally employ flaps, slots, or slats to augment the wing lift during takeoff and landing.

The area, S, is the area of the lifting surface. For example, an airplane with rectangular-shaped wings has a wing area equal to the wing span times the width of the wing.

$$S = 6 \times 36 = 216 \text{ ft}^2$$

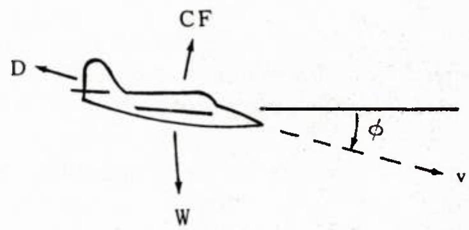


Glide Analysis

There are many approaches which can be used to analyze the gliding flight of a vehicle; most textbooks use theoretical mathematical approaches. However, in this text a less complex approach will be used to provide familiarization with the factors that influence the ground range of lifting vehicles.

The glide will be analyzed in three steps: (1) The Ballistic Period. The portion of the flight down to 250,000 ft is considered as ballistic (non-lifting) flight. In this case, 250,000 ft is defined as the atmospheric entry altitude. (2) The Altitude Hold Period. As the vehicle enters the 250,000 ft-level, the pilot will maintain altitude until the airspeed diminishes to the best glide speed. (3) The Descending Glide Period. As the vehicle approaches the best glide speed, a descending glide is initiated. The pilot maintains the glider at the attitude that corresponds to $\left(\frac{L}{D}\right)_{max}$ which is the flight attitude that will result in a minimum glide angle and maximum horizontal distance. The terminal maneuvers necessary to recover the vehicle are not considered part of the glide phase and will not be discussed.

STEP 1: THE BALLISTIC PERIOD



The lifting vehicle enters the earth's atmosphere at near orbital speed, and at a shallow angle. This attitude remains constant until the vehicle descends to approximately 250,000 feet. At this point the atmosphere has sufficient density to support lift at reentry velocities.

STEP 2: THE ALTITUDE HOLD PERIOD.—The pilot maintains altitude by gradually increasing the angle of attack as the speed decreases to the best glide speed. The *distance* travelled while the vehicle slows to optimum glide speed then becomes an important factor. Distance is the average speed multiplied by the time (v_{avg} x t). Speed will be examined first.

The external forces at the point of optimum glide speed are shown in Figure 10.

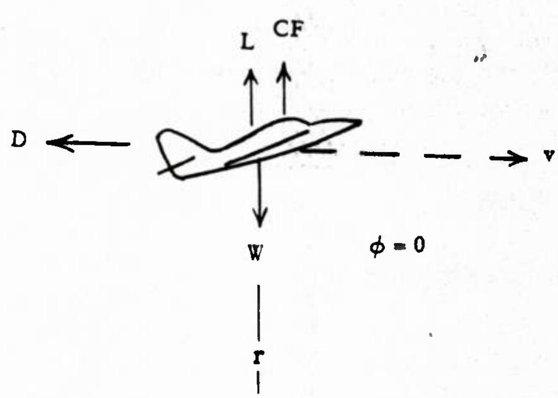


Figure 10. External forces at the point of optimum glide speed.

$$W = L + CF = \frac{C_L \rho v^2 S}{2} + \frac{Mv^2}{r}; \quad M = \frac{W}{g}; \quad CF = \frac{Mv^2}{r}$$

In order to obtain the average speed it is first necessary to solve the above relation for optimum glide speed, v_g, when C_L for optimum glide is used.

$$v_g = \sqrt{\frac{W}{\frac{C_L \rho S}{2} + \frac{W}{gr}}} \qquad (4)$$

C_L is obtained from the performance curves for the particular vehicle. The average speed, v_{avg}, is the entry speed, v_e, plus optimum glide speed, v_g, divided by 2, $\left(v_{avg} = \frac{v_e + v_g}{2}\right)$.

The next calculation in Step 2 is to find the time required to slow the vehicle to optimum glide speed while maintaining altitude. To do this, use Newton's Second Law of Motion, in the horizontal direction.

$$F = Ma, \text{ from the preceding figure}$$

$$F = D = \frac{C_D \rho \; v^2 S}{2}$$

$$M = \frac{W}{g}$$

$$a = \frac{\Delta v}{\Delta t}$$

Substituting into the latter equation and solving for Δt (the time required to slow to optimum glide speed),

$$\Delta t = \frac{2 \Delta v \; W}{g \, \rho \, v^2_{avg} \, C_D S}; \; \Delta v = v_e - v_g$$

$$\text{Distance} = (\Delta t) \, (v_{avg}) = \frac{2 \Delta v \; v_{avg} \; W}{g \, \rho \, v^2_{avg} \, C_D S} = \frac{2 \, \Delta v \, W}{g \, \rho \, v_{avg} \, C_D S} \qquad (5)$$

STEP 3: THE DESCENDING GLIDE PERIOD.—The solution for the glide angle, ϕ, can be resolved by analyzing the following equations, established at the beginning of this section:

(a) $L + CF = W \cos \phi$
(b) $D = W \sin \phi$

Dividing equation (b) by equation (a),

$$\frac{D}{L + CF} = \frac{W \sin\phi}{W \cos\phi} = \tan \phi$$

The magnitude of the centrifugal force is very small at glide speed and CF can be neglected. This leaves the solution of the glide angle in a simple form:

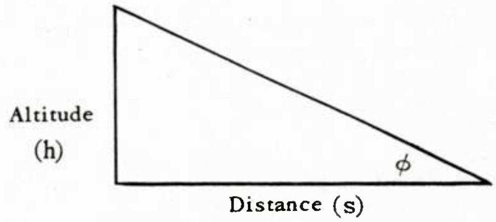
$$\tan \phi = \frac{D}{L}; \; \left(\text{or } \cot \phi = \frac{L}{D} \right) \qquad (6)$$

This equation indicates that once the $\frac{L}{D}$ characteristics of a vehicle are known, the optimum glide angle of the vehicle can be approximated. When the glide angle and altitude are known, the distance travelled in the glide is a simple relationship of the angle:

$$\cot \phi = \frac{S}{h}$$

$$S = h \cot \phi$$

$$S = h \left(\frac{L}{D}\right) \qquad (7)$$



Conclusion

This chapter presents some of the problems that arise during penetration of the atmosphere with emphasis on the heating and deceleration aspects. Only that detail has been presented which is needed to provide some understanding of the problem and to lay a foundation for further analysis. The problems of atmospheric penetration and reentry are at present under intensive study by space scientists and engineers.

BIBLIOGRAPHY

DUNCAN, ROBERT C. *Dynamics of Atmospheric Entry.* New York: McGraw-Hill, 1962.

ELY, LAWRENCE D. *Return from Space.* Springfield, Ill.: Charles C. Thomas, 1966.

GAZLEY, CARL, Jr. "Deceleration and Heating of a Body Entering a Planetary Atmosphere from Space" in Air Force. Office of Scientific Research. *Astronautics Symposium.* 1st— 1957— (Volume title: *Vistas in Astronautics)* New York: Pergamon Press, Symposium Publications Division, 1958.

LA GALLEY, DONALD P. *Ballistic Missile and Space Technology Re-entry and Vehicle Design, Vol. IV.* New York: Academic Press, 1960.

LOH, W.T.H. *Dynamics and Thermodynamics of Planetary Entry.* Englewood Cliffs, N.J.: Prentice-Hall, 1963.

MORROW, C. T., ELY, L. D., and SMITH, M. R. *Ballistic Missile and Aerospace Technology Re-Entry.* New York: Academic Press, 1961.

RAND Corporation. *Aerodynamic Aspects of Future Space Vehicles and Implications for Testing.* Washington: RAND Corporation, 1967.

————. *Atmospheric Entry of Manned Vehicles.* Santa Monica, Cal.: RAND Corporation, 1960.

————. *An Introduction to Astronautics, Vol. I.* Santa Monica, Cal.: RAND Corporation, 1958. Pp. 10.1 through 10.32.

TRUITT, ROBERT W. *Fundamentals of Aerodynamic Heating.* New York: Ronald Press, 1960.

CHAPTER 9

COMPUTERS

OVER THE YEARS man has dreamed of travel through space, but only recently has he been able to achieve it. Actually, space travel was not delayed until the present generation because of lack of knowledge of orbits and trajectories or a failure to understand the basic principles of physics. It was practical engineering, rather than a lack of knowledge or incentive, that delayed space travel. In recent years, two major engineering developments have made it possible to turn man's dreams about space exploration into a reality. These are the large booster rocket and the high-speed electronic computer. Until the development of the large booster rocket, man simply could not generate enough thrust to make space experimentation and exploration a reality. Until the modern computer was put into operation, man could not make all the complex computations with the speed and accuracy necessary to design space systems and then to guide and control their operations.

Today, computers are a part of almost every facet of space operation. They are essential to guidance and control; calculation of trajectories and orbits; tracking and cataloging space objects; prediction of times and places of reentry for space objects returning to the earth; managing the logistics of space operations facilities which support both research and development of space vehicles and space boosters; and operating simulators of the space environment in order to train astronauts. Computers operate reliably and quickly enough to be of great value in such areas as weather analysis, communications systems, command systems, logistics, personnel, accounting, and almost all other support systems essential to space operations. This chapter explains the elementary principles of computer operations, with special emphasis on the digital computer.

CLASSES OF COMPUTERS

Of the two principal classes of computing machines, analog and digital, the digital is the more versatile and the more widely used. However, analog computers operating alone or through a direct couple to a digital system perform important service in support of the space program and, therefore, will be considered briefly.

Operation of an Analog Computer

As the name implies, the analog computer operates by substituting physical analogs for numbers in computations and through variation of the physical

entities to achieve a computational result. Thus, in the operation of a slide rule, the logarithms of numbers are represented along a calibrated scale. Multiplication and division are then effected by adding or subtracting appropriate scale segments, an operation performed by physically moving two suitably calibrated scales relative to each other. Similarly, quantities may be represented through the angular deflection of a shaft or the response of an electrical circuit to changes in current or voltage.

Regardless of the analogy used in constructing an analog system, however, two properties of this type of computer are of particular importance as the system is applied in support of space operations. First, the analog computer deals with continuous functions and thus is more capable of representing physical processes than is a digital computer. This advantage, however, is mitigated by the fact that the degree of accuracy achievable by an analog system is limited by the tolerances to which it can be constructed and the ability of the components to remain stable with changes in environmental conditions and the passage of time.

Secondly, the analog computer has the ability to directly sense changes in physical systems without the need for an intermediate encoding device. Thus, in the case of a spacecraft, for example, an analog computer can be coupled directly to the output unit of an electrical sensor and perform a series of tasks based upon changes in the output voltage from the device to which it is attached.

Application of Analog Computers in the Space Program

Analog computers can be applied to a variety of problems connected with the space program. For example, an analog system can be connected to a tracking antenna in such a way that the output of the computer determines the azimuth and elevation of the antenna. Further, tracking data from the antenna can be used as input to the computer. The closed system thus formed makes it possible to automatically follow any specified target with the antenna for the purpose of either gathering data from or transmitting data to a specified vehicle.

Analog computers have been placed on board space vehicles for the purpose of guidance and attitude control. In these cases the computers have used inputs generated by stable platforms or arrays of sensors and have manipulated the attitude control systems with the analog output. As in the case of the antenna tracking problem cited previously, the computer and the vehicle control system form a closed command loop.

Hybrid Computation

While purely analog systems of the type described above are useful in many applications, there are situations in which it becomes desirable to take advantage of the increased accuracy and computational flexibility of the digital computer. Yet, the data produced by a moving antenna or spacecraft are essentially analog in nature and must be digitized if they are to be processed by digital computer. Further, if the digital output is to be used in a command loop to physically position mechanical components, it must be converted to analog impulses acceptable to the control mechanisms involved.

To satisfy the requirement for performing the conversion from analog to digital data, hybrid computer systems have been constructed to incorporate the advantages of both analog and digital computers. The analog components can be used to sense or control measurable entities. On the other hand, the more versatile digital system is used to satisfy the requirements for computation and analysis.

Operation of Digital Computers

As the name implies, the digital computer operates with discrete entities or digits. As a result, the degree of accuracy with which it operates is directly proportional to the size of the registers it uses in performing the required operations. For example, a digital computer whose registers are capable of holding eight digits is more accurate than a machine whose registers will accommodate only four digits.

While an analog computer operates through a continuum, the digital computer operates through a series of discrete steps. In fact, the digital computer must be explicitly instructed to execute each step required for the solution of a problem, no matter how trivial the step may seem to be. For this reason, the programs or sets of instructions required by the machine before it will perform a task can become very long and complex. Hence, the time required to prepare a program may be quite considerable, and it becomes extremely difficult to change long programs once they have been written.

MAN-MACHINE COMMUNICATION

In order to minimize the time required to prepare and verify the accuracy of digital computer programs, a series of "problem-oriented"* languages have been developed. These languages are structured so that they closely resemble the notation used in the statement and solution of particular classes of problems. Thus, the solution of mathematical problems is often programmed in the FORTRAN (Formula Translation) language using a notation very similar to common algebra. On the other hand, COBOL (Common Business Oriented Language), a language used in the solution of management data processing problems, is structured in a form that closely resembles the format of instructions as they might appear in a formal office operating procedure.

Supplementing the "problem-oriented" languages that have simplified significantly the process of writing computer programs, are the libraries of general purpose programs that have been prepared and made available to all users at a particular installation. Thus, it is often not necessary for an individual to take the time and effort required to write a unique program to solve a specific problem. Further, it is not always necessary that he become familiar with one or more programming languages in order to obtain the benefit of computerized computational support.

To take advantage of the library programs, the user first selects from a catalog a general purpose program that will perform the operations he requires.

* Problem-oriented language—A source language oriented to the description of a particular class of problems.

If a suitable routine is found, he has only to put his data in the form specified and submit the entire package for processing.

If no program is found in the catalog that meets exactly the needs of the user, he may select one that could be modified with relative ease to perform the required tasks. While this course of action may save considerable time and effort in those cases where major modifications are not needed, two potential areas of difficulty must be recognized. First, in order to be able to attempt modification of an existing program, the user himself should be an experienced programmer. Secondly, it may take more time to modify a complex program than it does to write a new one.

Note that users can benefit from work done at a number of computer installations through the program exchange organizations that have been established by users of particular types of computers. Such organizations make available large numbers of general purpose programs, and indexes of available routines should be checked by all users before beginning the development of new program packages.

Thus, the problem of language in man-machine communication has, to some extent, been solved. In many cases, library programs are available. Should it be necessary to develop a program or set of programs for a particular application, the availability of "problem-oriented" languages greatly simplifies the tasks of preparation and check-out.

In addition to the "language barrier" that exists between man and machine mentioned in the preceding paragraphs, there is a basic incompatibility in the communications media used by each. Men understand written and oral symbols and sounds that may vary in structure to a great extent and still be intelligible. Written script, for example, is usually understandable in spite of the fact that handwriting varies greatly from one individual to another. Similarly, people can speak and be understood, although accent and idiomatic structure vary greatly from one geographic area to another.

The case with computers, however, is much different. Inputs to machines at this point in time must be highly structured. The media used (such as the punched card) may not be easily interpreted by man. As a result, a significant problem exists in translating material into a machine-processable form.

Progress is being made in overcoming these difficulties. Optical scanners capable of reading typewritten material are operational, and some progress is being made in the machine recognition of script. Photo analysis and interpretation using digitized optical inputs also are being developed. Further, research and development in the areas of machine processable audio inputs and machine generated audio responses is also in progress.

To facilitate thc input of information to computers, keytape and similar devices are replacing, to a degree, keypunches. Also, specialized terminal devices which sense card or key input are being installed to gather data at the point of origin. High-speed output through cathode-ray tube displays is coming into use. Also, devices that will generate at a very high speed images on photographic film for later viewing are under development and should soon be readily available.

While it is not yet possible to carry on a conversation with a computer as one would with another human being, work leading toward such a development has shown progress. It is not beyond the realm of possibility that such conversations will be a reality in the future.

DIGITAL COMPUTER CONTROL

The high speed at which the modern digital computer operates makes it impossible to exert manual control over the details of operation. The control problem is complicated further by the fact that some components of the machine operate at speeds a thousand or more times greater than others. For example, the computer may be able to retrieve information from a high speed storage unit in about a millionth of a second, but it takes a printer more than a hundredth of a second to transfer the information to paper, a speed differential of ten thousand times. Lastly, because of the cost of modern digital computers and the existing demands being made upon available time, it is important to have a control system that will insure maximum utilization of the available equipment.

The Role of the Monitor

In order to effect the required control, computer programs have been written to supervise or monitor the operations of the machine. The monitor serves: (1) to bring user programs into the machine for processing; (2) to insure that all of the programs from the system's library required by the user's program are made available at the appropriate time; (3) to turn control of the machine over to the user's program to perform the required tasks; (4) to make available to the user's program those auxiliary devices such as tape drives and line-printers required during execution; and (5) to transfer control of the machine to the next user's program when a task has been completed. The monitor also provides the operators with the information they need to manage the computer system efficiently. Thus, the monitor is actually an automated operator for the computer, and it controls the machine so that little, if any, human intervention is required.

BATCH PROCESSING OPERATIONS

In general, computer systems may be identified as operating in either a "batch-processing"* or a "time-sharing"** environment. Within each of these classes operations may be geared either to provide a "real-time" response to requests for computer service, or a considerable lapse of time may be permitted between the time a job is submitted for processing on a computer system and the time that the output is returned to the user.

* Batch processing—A technique by which items to be processed must be coded and collected into groups prior to processing.

** Time sharing—The apportionment of intervals of time availability of various items of equipment to complete the performance of several tasks by interlacing.

The Monitor and the "Batch-Processing" Environment

When a digital computer is operated in the batch-processing mode, the need for a monitor is minimal. In fact, under some conditions, it is possible to operate a batch-processing computer system with a reasonable degree of efficiency without the use of a monitor program to provide automatic control. However, the installation of a monitor will almost always improve the efficiency of computer operations.

Within the batch processing mode, the monitor may have little to do except provide the library and system programs that a particular job requires and to take care of the "housekeeping" chores after the job has been completed. For example, if a programmer submits a FORTRAN program for compilation and execution, the monitor must, after sensing that a job is to be processed, call in the FORTRAN compiler and give it control of the machine. Once the process of translating the FORTRAN program to machine language has been accomplished and the compiler has obtained from the system library those programs and subprograms that are required, the monitor transfers control to the compiled FORTRAN program, and execution takes place. Then, when the program has run its course, control of the computer is again transferred to the monitor, and it begins the cycle for the next job that must be processed.

Many monitors have under their direct control several subsystems and/or compilers. Thus, the systems monitor must sense which of the several subsystems must be given control of the computer in order to satisfy each requirement and, in addition, set up the instructions sequence to regain control from the subsystem once a job has been completed. The hierarchy of subsystems operating under a given monitor may be several layers deep. For example, the overall operating system may have under it a monitor that generates reports and another that supervises the various compilers that may be operating on the machine. Then, under control of the latter, one may find FORTRAN and COBOL compilers and an assembly language processor.

Monitors operating in a batch processing mode should also have some means of recovering from machine or program errors without requiring intervention by a human operator. In this way the continuity of operation is preserved, and a minimum time is lost because of failures. Unfortunately, monitors do not recover automatically from all failures. As a result, the need for operator intervention has not been eliminated completely.

The more sophisticated batch-processing systems may be designed to store operational programs in the systems library and to recall them as needed for data processing. In this case, the operating system must be able to obtain the required programs from the library, give them control of the computer, and regain control once the job has been completed. Operating systems of this type also must be able to add programs to the library, delete those that are no longer needed, and make changes to stored programs as they are required. Lastly, the most effective operating system would adjust the position of programs stored in the library so that those that are most frequently used are the most accessible.

"Real-Time" Batch-Processing Operations

During many phases of space operations, command and control are better handled by computers than men because of the requirement for rapid reduction of data and a short response time. In this environment, the computers performing the command and control function must receive data from one or more sources, analyze it, and output the results in the form of a command to a vehicle or of an information display to a human flight controller. When operating in this mode, the computer is said to be operating in "real time."*

A batch-processing computer operating to satisfy a "real-time" requirement is dedicated to the processing of a single program during the period when it is performing the command and control function. However, the program may be written so that it senses and processes data from several sources, and then it produces output through several channels that are used by both ground support and airborne systems. The "program" in this case would actually be segmented to form a system of related routines that are processed on an "as required" basis under the control of a supervisory routine that guides the overall operation of the machine. Nevertheless, only one program or program segment would be operational on the computer at any one time. Therefore, a system operating in this environment may be identified as a "real-time, batch-processing" system.

TIME-SHARING COMPUTER OPERATIONS

As mentioned previously, the various electro-mechanical devices of a computer operate at least one thousand times more slowly than the central processor and the high speed storage unit. In the batch-processing mode, the entire computer must remain idle for a period of time after a command has been given to an electro-mechanical device, such as a printer, while that action is completed. During this period, a tenth of a second or more, the computer could be performing several thousand computational operations. Thus, when a program requires substantial time for input and output operations, a majority of the total time the central processor is available to perform computation may be spent waiting for mechanical operations to be completed. Any technique that would permit the central processing unit to perform some other task while waiting for a mechanical operation to be completed would increase effectively the utilization of the computer system.

In practical operation, the technique used is computer "time sharing." With this technique, two or more programs are actually in an operational status at the same time. Control of the computer is transferred from one program to another on the basis of predefined criteria. Thus, when one of the programs requires that a line be printed, the command to the printer is given, and control of the central processor is transferred to another program. Execution of the latter program is resumed at the point where it was last interrupted; it is then permitted to continue for a predetermined period, or until it too requires that an input or output operation be performed. Although only one program at a time

* Real time—Pertaining to the performance of a computation during a related physical process to obtain results needed to guide that process.

is actually being processed, the central processing unit is being kept active. Therefore, the capability of the machine is more fully utilized when it operates on a "time-sharing" basis.

Because there is an overhead of computer "housekeeping" in "time sharing" when control is transferred from one program to another, this mode of operation is not suitable for processing programs that require a great deal of computation and few input or output operations. Thus, if the machine is being shared by a computationally oriented program with nine others regardless of type, it will receive only about one tenth of the available computer time. Under these conditions a program that would take one hour if allowed to run to completion would require ten or more hours on a "time-shared" system.

"Time sharing," therefore, is not the universal cure for all computing problems. Where there is significant input and output operation or interaction with a slowly responding environment, "time-sharing" can increase significantly the efficiency of central processor utilization. On the other hand, where there must be a great deal of computation accompanied by few input or output operations, the batch-processing mode is preferable.

Operation of the "Time-Sharing" Monitor

Unlike the batch-processing operation, "time sharing" requires a suitable monitor to supervise the computer through program control. It would be virtually impossible for a human being to perform the monitor tasks because of the required reaction speed.

Not only must the "time-sharing" monitor perform all of the batch-processing monitor tasks, but also is must do much more. First, it must be designed to distribute the available work load so that all of the existing computational requirements are met. Second, the monitor must be able to perform all of the "housekeeping" operations of transferring control of the machine from one program to another. Third, since there is no way that the programmer can identify the physical units required by his program and be sure he is not in conflict with another simultaneous user, the monitor must handle the assignment of input/output and auxiliary storage devices. Last, the monitor has to check on the physical status of the computer and insure that no program interferes with another. This latter feature is not required of the monitors of some of the newer machines because they have been designed to perform these functions automatically.

Time Sharing in Support of Space Operations

"Time-sharing" computer systems are, by their very nature, not particularly sensitive to the rate of input or output. They can accept both high rate telemetry data or very slow rate manual teletype data without necessarily degrading the system performance. Hence, machines operating in a "time-sharing" mode are particularly suitable for operations that require interaction between the computer and man, or the computer and devices requiring only intermittent

service. Such operation can produce limited information rapidly, eliminating the need for the receiver to have at its disposal large amounts of data for which there may be no requirement.

RELIABILITY OF COMPUTER OPERATIONS

Because of the critical role computers often play in support of space operations, it is essential that every precaution be taken to insure that they operate reliably. If this is not done, the result could again be, as it has in the past, the complete failure of a mission with the attendant financial loss and slippage in the program of which the mission was a part.

Two types of failure must be prevented: mechanical failure of the machine and program error. The former must be compensated for at the time of occurrence; and the probability of failure may be reduced by an aggressive schedule of preventive maintenance. Program error is the result of human error. It may develop from failure to build the computer program to account for an event which occurs during the mission, resulting in partial or total failure. Another possibility for error may occur even if all contingencies have been allowed for. In such a case, check-out of the program with available test data may not have been carried out deeply enough to cause failure during the test phase. Finally, the programs themselves must be carefully loaded on the machine and properly initialized.

Testing of computer programs which deal with space flights is critical. Many different events can take place during a space flight, some of which are virtually unpredictable. The computer engineer must try to plan for these contingencies when he designs the testing procedure. In addition, the programs must check themselves and then either recover from error automatically or initiate a signal which lets flight controllers know that something is wrong.

Mechanical Failure of Computer Systems

Mechanical failure of a computer is somewhat easier to deal with once a procedure to detect errors has been established. So long as failure is obvious to the operators of the system, corrective action can be taken. However, there are times when mechanical failure is extremely difficult to detect by inspection of the output of the machine or through monitoring of the operator's console. Yet, these small errors can be disastrous in terms of mission accomplishment.

Detection of machine failure can be facilitated by operating two or more identical computer systems in parallel. Each system receives identical inputs but is completely independent of the others in operation. Discrepancies in the outputs of two systems can be detected quickly, and steps can be taken to identify the machine that has failed. Fortunately, many mechanical failures are recognized easily by human controllers. The faulty machine can then be removed from service. In cases where the discrepancy in output between two machines is a relatively small one, predefined test procedures must be used to identify and correct the cause of failure. Once the source of the error has been identified, the computer which is operating properly is given control of the mission

while the other is shut down, repaired, and returned to function as a backup in case of further failure. The benefits of this type of operation were realized during the reentry phase of the Glenn flight in the Mercury program.

The stress resulting from long periods of operation without maintenance may cause mechanical failure of computers. Hence, time for preventive maintenance must be allowed in the operating schedule. The only way in which this can be accomplished while maintaining an independent parallel backup is to have independent systems available to satisfy each requirement for mission support. In this arrangement, one machine is active, a second is providing backup through independent parallel operations, and others are undergoing preventive maintenance or providing normal batch processing support. Three IBM 7094 systems were used in this way to support the fourteen-day Gemini mission. Five IBM 360/75 systems are used to support Apollo.

DATA COMMUNICATION

One of the keys to the success of the space program is the integration of data streams gathered from widely separated tracking stations and from vehicles in flight into an organized information flow to be used in mission command and control. The data communications system capable of satisfying this requirement must have high capacity for rapid and reliable data transmission, and be able to format and channel this data to obtain the desired organization.

Thus, advances in data communications technology have contributed materially to the progress of space programs. When the Mercury program began, instantaneous transmission of data over short distances was not reliable. Further, it was not considered possible to transmit data reliably from tracking stations in Australia to Goddard Space Flight Center, Greenbelt, Maryland, in less than two hours. Presently, with advances that have occurred in communications technology, it is possible to organize and transmit data reliably at a high rate (hundreds of thousands of bits per second) over long distances. In the Apollo system data is formatted and channeled by three Univac 494 computers.

Data Transmission for Command and Control

The data transmission problem can be divided into two major categories. The first of these is the transmission of data to and from vehicles in space for the purpose of command and control. This must be done both quickly and efficiently. Further, it may be necessary not only to communicate directly with the vehicle and process the data rapidly, but also to assemble the various data streams at a central location for integrated processing and retransmittal. Thus, a significant amount of surface-to-surface data communications, in addition to surface-to-vehicle data communications, may be involved.

Further, the results of processing may have to be transmitted over a substantial distance from the computer installation to a tracking station that is in position to communicate with the vehicle. Again, the capability to transmit at high speed over substantial distances may be required.

Recording of Experimental Data for Later Processing

The second data transmission problem is that of receiving information from a vehicle for processing at a later time. In this case, the problem is somewhat simpler than the command and control problem because the data need only be recorded on site rather than being retransmitted immediately to a centralized computer for "real-time" processing. Thus, the data communications problem is confined to transfer between the vehicle and the surface and to recording the data in a form suitable for processing and analysis at a later date.

SELECTIVITY IN THE RECORDING OF EXPERIMENTAL DATA.—Since the volume of data gathered by vehicles and transmitted to ground stations can become very large, the question of selectivity of recording must be considered. For example, if unchanging voltmeter readings are transmitted to the surface continuously, much virtually useless data has been recorded. However, if only significant changes in readings are recorded, the information available for later analysis would be of greater significance. Thus, computers used to receive and record data may be used to perform preliminary edits that would eliminate much processing at a later date and yet provide researchers with a significant record of vehicle operation.

COMPUTERS IN SUPPORT OF SPACE OPERATIONS

Because of their versatility, digital computers operate in support of virtually every aspect of space programs from the basic design of boosters and payloads to the reduction and analysis of the data after mission completion. Computers function to schedule experiments for various vehicles, to keep managers and technicians informed of the latest status of a vehicle as it moves along the production line, and to position vehicles to a desired attitude and trajectory.

One system, for example, uses a "time-shared" computer to give the engineers and technicians access to the latest data on vehicles at various stages of production. Using this data, it is possible to recompute the expected performance figures after modifications to the original plan have been made and to decide what further design changes will be required. The data base is carefully controlled by the project office, and only approved changes are permitted. Yet, all personnel with access to the system can manipulate the data base in any way they wish so long as no permanent changes are made in its structure.

Computers in Active Support of Flight Operations

Digital computers perform several important functions during prelaunch and launch activities. They are used to monitor the status of the vehicle as well as that of the tracking network. They operate during the launch phase to control the flight of the vehicle to injection into an orbit or a deep space trajectory. If an abort is necessary, the procedures are controlled by the computer.

Computers also calculate the deorbit point and time for the reentry maneuver. Insofar as possible, they are used in monitoring the letdown sequence.

On-Board Computer Systems

The Apollo vehicle has an on-board digital computer which is used for flight control and navigation. Data may be entered into this system through a key-

board on the console, and readouts are presented in the form of digital displays. Benefits of these systems to this program have been widespread. During Gemini, the on-board system on one of the vehicles was used to override the ground support system during reentry.

SUMMARY

This chapter has presented some of the fundamentals of computer operations and the way computers function in support of the space program. Like other more publicized subsystems, we would not have a national space program without the availability of a high-speed and accurate computer capability.

BIBLIOGRAPHY

Aerospace Defense Command. *Delta Program Documentation.* 1967 and 1968.

ASELTINE, JOHN A., ed. *Peaceful Uses of Automation in Outer Space.* New York: Plenum Press, 1966.

BROWN, KENNETH and WEISSER, PETER, eds. *Ground Support Systems for Missiles and Space Vehicles.* New York: McGraw-Hill, 1961.

CROWLEY, THOMAS H. *Understanding Computers.* New York: McGraw-Hill, 1967.

DESMONDE, WILLIAM H. *Computers and Their Uses.* Englewood Cliffs, N.J.: Prentice-Hall, 1964.

KLERER, MELVIN and KORN, GRANINO A. *Digital Computer User's Handbook.* New York: McGraw-Hill, 1967.

MARTIN, JAMES. *Programming Real-Time Computer Systems.* Englewood Cliffs, N.J.: Prentice-Hall, 1965.

CHAPTER 10

RELIABILITY OF SPACE SYSTEMS

THE DEVELOPMENT of the digital computer has made possible a high degree of accuracy in computations for space research and engineering and has increased overall efficiency. In the operation of any space system, accuracy and efficiency are paramount. The same is true for any ballistic missile system. Since these systems are all very expensive, operational failure is extremely costly. In addition, the success of the system is nearly always very important not only to the Air Force but also to the nation as a whole. Each space shot carries with it the prestige of the United States, and the very survival of the free world may rest upon the reliability of missile systems.

The advent of missiles and space systems has outmoded the philosophy of "fly and fix." Good systems must operate when fired; "holds" and mission aborts must be kept to a minimum. To accomplish this end, both manufacturers and operators of space systems must pay more attention than ever before to the reliability of the system as a whole. This requirement poses two critical problems: First, how can reliability be measured? Second, how can it be increased? Both of these problems are discussed in this chapter, but, as a prelude to their discussion, it is necessary to study the subject of probability, at least in its simplest form. *Realibility is a term meaning the probability that equipment will perform a required function under specified conditions, without failure, for a specified period of time.* Notice that reliability is a probability. It is a probability related to a system under specified conditions and for a specified time.

PROBABILITY

The probability that a given event will occur is defined as the number of ways in which the event can occur, divided by the number of ways in which the event can occur, *plus* the number of ways in which the event can fail to occur.

If N_s is the number of ways in which a given event, A, can occur and N_f is the number of ways in which the event can fail to occur, then the probability of occurrence of the event is:

$$P(A) = N_s/(N_s + N_f) \qquad (1)$$

The probability of the event not occurring is:

$$P(\text{not } A) = N_f/N_s + N_f) \qquad (2)$$

It is obvious from the definition of probability that values of probability range from 0 to 1. A 0 probability represents an impossibility, whereas a probability of 1 represents an absolute certainty.

Another important relationship can be found from the definition of probability by adding equations (1) and (2) as follows:

$$P(A) + P(\text{not } A) = 1 \tag{3}$$

Two conditions must be met, however, before the three equations given above may be used: (1) The number of possible outcomes in a particular event must be known. (2) The possible outcomes which are favorable (or unfavorable) to the event must be known.

Probability can be illustrated by rolling a pair of unbiased dice. For example, what is the probability of obtaining a 7 in one roll? To answer this question, consider one roll as an event and then count the number of ways that the event can result in rolling a 7. The possible ways to succeed, then, are to obtain the following combinations: 1 and 6, 2 and 5, 4 and 3, 3 and 4, 5 and 2, and 6 and 1; that is, there are six ways of succeeding. The total number of ways that the event can occur is 36, and the number of ways in which the event can occur and fail is 30. Therefore:

$$P_s = \frac{N_s}{N_s + N_f} = \frac{6}{6 + 30} = \frac{6}{36} = \frac{1}{6}$$

$$P_f = \frac{N_f}{N_s + N_f} = \frac{30}{6 + 30} = \frac{30}{36} = \frac{5}{6}$$

In situations dealing with a limited number of items, such as dice, cards and coins, it is relatively simple to count the possible events and meet the above conditions. In many instances, however, the number of possible outcomes is too great to count, or the final result will depend upon the favorable or unfavorable outcomes of more than one event. In cases such as these, other equations and theorems must be used to calculate probabilities.

Mutually Exclusive Events

Two or more events are said to be mutually exclusive if one and only one event can occur at one time. In a mutually exclusive sequence, the probability that either one or another of the events will occur is equal to the sum of the probabilities of occurrence of the individual events. A mathematical statement of this rule can be made as follows:

$$P(A \text{ or } B) = P(A) + P(B) \tag{4}$$

This equation can be extended to include any number of mutually exclusive events. The application of this rule can best be shown by a simple problem.

Consider a box containing 1 white ball, 1 red ball, and 3 black balls. What is the probability of drawing a red or white ball in one draw? Successful accomplishment

of the endeavor can be effected in two ways: drawing a red ball or drawing a white ball, only one of which can be drawn on any one trial. This is a case of mutually exclusive events; consequently, the probability is the sum of the probabilities of drawing a red ball and drawing a white ball.

From equation (1), these are 1/5 in each case. Therefore the probability of drawing either a red ball or a white ball is by equation (4):

$$P(\text{red or white}) = 1/5 + 1/5 = 2/5$$

The probability of drawing a black ball can be found as follows:

$$\begin{aligned} P(\text{red or white}) + P(\text{black}) &= 1 \\ P(\text{black}) &= 1 - 2/5 = 3/5 \end{aligned}$$

Contingent Probabilities

Suppose the undertaking is some large endeavor consisting of a number of separate events. If success or failure of the endeavor depends upon success or failure of each of the events in the endeavor, the probability of the endeavor is contingent upon the probability of the events. If this is the case, the probabilities for each of the events are combined by multiplying them, not by adding them. The probability of success or failure for the endeavor is obtained by multiplying the probabilities of success or failure for each of the events in the endeavor. Suppose the endeavor is to launch a Saturn booster. Success or failure in this endeavor is contingent upon the success or failure of a very large number of events, but the example will be simplified by discussing success of first stage performance. If the probability of success in fueling is .9 and the probability of successful ignition is .8, what is the probability of success in the endeavor? Since success in the endeavor is contingent upon success in both events, the answer is computed by multiplying the probabilities of success for the two events. Thus, probability of a successful launch is equal to .9 times .8 or .72.

Sometimes, success in an endeavor may be contingent upon failure in one or more of the component events. For example, to succeed in getting from the base to the bomb release line, an aircrew must fail to abort, fail to be shot down, and fail to make a gross navigational error. In some cases, success in the endeavor is contingent upon success in some events and failure in other events.

Events may be further classified as being either independent or dependent. Independent events are those in which the outcome of one event does not affect the outcome of any other event in the endeavor. Dependent events are those in which the outcome of one event does influence the outcome of other events in the endeavor.

EXAMPLE I—CONTINGENT INDEPENDENT (successful occurrence).—What is the probability of drawing 2 black balls from a box containing 3 black balls and 2 white balls, if two draws are made and the ball drawn is replaced prior to the second draw? In this instance, the endeavor consists of two events, and the successful outcome of the endeavor is contingent upon being successful in each event, that is, in drawing a black ball each time. The events are independent in that the first draw will

not affect the second draw, since the ball will be replaced. The probability of drawing 2 black balls is the product of the probabilities of drawing a black ball each time. The probability of drawing a black ball is 3/5 in each instance. Therefore:

$$\begin{aligned} P(\text{2 black}) &= P(\text{black}) \times P(\text{black}) \\ &= 3/5 \times 3/5 \\ &= 9/25 \text{ or } 36\% \end{aligned}$$

EXAMPLE II—CONTINGENT DEPENDENT (successful occurrence).—What is the probability of drawing 2 black balls from the box if the first ball drawn is not returned to the box? This problem is similar to example I except that the events are dependent, since the probability on the second draw is dependent on the outcome of the first draw. The probability of getting a black ball on the second draw is either ¾ or ½, depending on whether or not a black ball was drawn the first time. The probability of drawing a black ball the first time is 3/5. If the first ball is black, then the probability that the second ball will be black is ½. The probability that both balls will be black is the product of these two probabilities:

$$\begin{aligned} P(\text{black then black}) &= 3/5 \times 1/2 \\ &= 3/10 \text{ or } 30\% \end{aligned}$$

Thus far, the examples have been such that the successful occurrence of an endeavor was contingent on the successful occurrence of the separate events. There are also endeavors in which the nonoccurrence, or failure, of an endeavor is contingent on the nonoccurrence, or failure, of the separate events. The following example will illustrate this.

EXAMPLE III—CONTINGENT DEPENDENT (success and failure).—What is the probability of drawing at least 1 black ball in two draws if the first ball drawn is not returned to the container? To draw at least 1 black ball one need not draw a black ball on each attempt. Rather, not drawing at least 1 black ball in the endeavor is contingent on not drawing a black ball on each attempt. Since the first ball is not replaced, the events are dependent. The problem can be solved as follows:

$$\begin{aligned} P(\text{no black then no black}) &= P(\text{no black}) \times P(\text{no black}) \\ &= 2/5 \times 1/4 \\ &= 1/10 \text{ or } 10\% \end{aligned}$$

If the probability of not getting a black ball is 1/10, then the probability of obtaining at least 1 black ball is 1 minus $1/10 = 9/10$.

EXAMPLE IV—CONTINGENT INDEPENDENT (success and failure).—What is the probability of drawing at least 1 black ball from the container in two draws if the first ball drawn is returned to the container? This example is similar to example III except that the events are now independent, since the first draw does not influence the second draw. For simplicity, success may be defined as the act of obtaining a black ball. Failure, then, is the inability to obtain a black ball and is denoted by P_f.

$$P(\text{no black, no black}) = P_{f_1} \times P_{f_2}$$

$$P(\text{no black, no black}) = 2/5 \times 2/5$$

$$P(\text{no black, no black}) = 4/25 \text{ or } 16\%$$

$$P(\text{at least one black}) = 1 - .16$$

$$= .84 \text{ or } 84\%$$

The above equations can be combined as follows:

$$P(\text{at least one black}) = 1 - P(\text{no black, no black})$$

$$= 1 - (P_{f_1} \times P_{f_2})$$

Since P_{f_1} equals P_{f_2}, this can be written as:

$$P(\text{at least one black}) = 1 - (P_f)^2$$

And for the general case, the equation is written as:

$$P(\text{at least one A}) = 1 - P(\text{no A})^n \qquad (5)$$

Equation (5) gives the probability of *at least one* occurrence in a series of n repeated trials. The above equation can be used to calculate the number of trials necessary to achieve a given probability of at least one occurrence. In space systems, as explained later, it can be used to calculate the reliability of redundant systems.

EXAMPLE V—MINIMUM PROBABILITY.—How many draws must be made from a container of 3 black balls and 2 white balls to have a minimum probability of 90% of obtaining at least 1 black ball? The problem can be solved by using equation (5).

$$P\ (\text{at least one black}) = 1 - P(\text{no black})^n$$

$$.90 = 1 - (.4)^n$$

$$-1 + .90 = -(.4)^n$$

$$.10 = (.4)^n$$

$$n = 2.51$$

Therefore, it would take three draws to have a minimum probability of 90% of drawing at least 1 black ball.

A Guide for the Solution of Probability Problems

Step 1. Define "success" for the problem under consideration.

Step 2. Does success in the endeavor depend upon the outcome of a series of events?

a. If "no," compute the single probability and the problem is solved.

b. If "yes," go to Step 3.

Step 3. Are these events mutually exclusive? Does being successful in one way *rule out* being successful in any other way?

a. If "yes," determine the probabilities of success for the several mutually exclusive events that may happen. The *sum* of the probabilities of success of the several events equals the probability of success in the endeavor, and the problem is solved.

b. If "no," go to Step 4.

Step 4. Does success in the endeavor require success in each event?

a. If "yes," determine the probability of success in each event. The *product* of the probabilities of success in each event is equal to the probability of success in the endeavor, and the problem is solved.

b. If "no," go to Step 5.

Step 5. Does success in the endeavor require "nonsuccess" in each event?

a. If "yes," determine the probability of failure in each event. The *product* of the probabilities of failure in each event is equal to the probability of success in the endeavor, and the problem is solved.

b. If "no," go to Step 6.

Step 6. Does success in the endeavor require success in some of the events and failure in others?

a. If "yes," determine the probabilities of success for those events where success in the endeavor depends upon success in the events; determine the probabilities of failure for those events where success in the endeavor depends upon failure in the events. The product of these probabilities is the probability of success in the endeavor, and the problem is solved.

b. If "no," then failure in the endeavor must be contingent upon failure in each of the events. It is then necessary to determine the probability of failure in each event. The *product* of the probabilities of failure in the events is equal to the probability of failure in the endeavor. The corresponding probability of success can be found by using the following formula:

$$P_s = 1 - P_f$$

(Where P_s = Probability of success in the endeavor

and P_f = Probability of failure in the endeavor.)

RELIABILITY

Suppose a booster for a spacecraft is fueled and checked, and the countdown is complete. All systems are "go." Will the booster and its spacecraft actually perform their mission without failure for the specified period of time? Actual experience shows that, under these conditions, some do, and some do not. However, before the actual launch; the directors of a project want a good estimate of the chances of success, or of the reliability of the system.

Suppose a booster is in the process of design, development, and production. With the thousands of component parts that must be assembled into an operating system, some mistakes inevitably will be made both in design and in the process of production to meet the design specifications. What can be done to measure the reliability of a design and of the component parts that enter into the finished product? Reliability, which is a word that is becoming more and more associated with both space systems and missile systems, is a probability idea. As stated earlier, it is the probability that a system will perform a required function under specified conditions, without failure, for a specified period of time. This idea can be applied to a complex system consisting of a multistage booster, a spacecraft with more than one stage, and a recovery system; it can be applied to one stage of a system; or it can even be applied to one component part, such as a transistor or a valve.

Calculating the Reliability of a System from the Reliability of its Parts

Even though a system may have as many as 30,000 parts, assume a simple system consisting of 10 parts for purposes of illustration. Also, assume that the reliability of each part has been measured and found to be .90. If the system is to operate, each of the parts must operate; therefore, the probabilities of success are contingent. What is the reliability of this simple system? Clearly it is $(.90)^{10} = .35$.

At first glance, .90 seems like a rather good reliability. But if the reliability is no greater, even a 10-component system becomes much too unreliable. A 30,000-component system would have a reliability so low as to be ridiculous. What can be done? First, the component parts of a system must have reliabilities much higher than .90.

Assume a 10-component system in which each part has a reliability of .99. Now the reliability of the simple system is $(.99)^{10} = .905$. Suppose the reliability of the parts is .999. Now the reliability of the 10-part system becomes $(.999)^{10} = .991$.

Now think of a system with 10,000 parts. One important conclusion from the simple example is immediately apparent. If reliability in a real system is to be something like .90, the reliability of the parts must be very high indeed. All of the parts must be designed and tested, redesigned and retested, until their reliability approaches 100%. In a recent symposium on reliability, engineers talked about the reliability of parts in the order of .9999 and higher.

METHODS FOR MEASURING.—Reliability of equipment is sometimes computed by first measuring the failure rate. For example, a part for a booster might be tested by selecting a number of the parts at random and starting them to operate under conditions which approximated the actual conditions of expected operation as closely as possible. Suppose that, on the average, one part failed each 100 hours of operation. Then, the failure rate, f, would be 0.01 per hour.

Another, and more common, procedure is to measure the mean time before failure (MTBF) and, from this, calculate the failure rate. For example, suppose the first part failed to operate after 40 hours, then more and more of the parts began to fail, but the last sample did not fail until 300 hours had passed. From this, the mean, or average, time of operation before failure can easily be calculated. Suppose it is 100 hours. Then $f = \frac{1}{\text{MTBF}} = \frac{1}{100 \text{ hours}} = .01$ per hour.

If the failure rate and the desired operating time of the part are known, the probability that each part will operate for this time can be calculated by using the formula: $P_s = e^{-ft}$ (where P_s is reliability, f is failure rate and t is time of desired operation). For example, if a piece of equipment had a failure rate of 0.001 per hour and was expected to operate for 10 hours, its reliability would be:

$$e^{-(0.001)(10)} = e^{-.01} \approx 0.99.$$

ADVANTAGES AND DISADVANTAGES OF THE PROCEDURE.—The procedure outlined above and numerous modifications of it are widely used by industrial firms. It has many advantages. In the first place, it enables the engineers to identify parts which have a reliability that is too low. They can then redesign the part and test it again and, through repetition of the procedure, improve the product. Further, it is not too complicated a procedure. It involves only careful testing and some calculation. Are the results, however, truly valid for testing the reliability of a complex system? Can this be done by testing individual components and calculating contingent probabilities?

The question focuses attention on some of the disadvantages of the system. The procedure assumes that the reliability of each component used in the calculation is constant. This may or may not be true. It is not true if the part has some minor modification made after the reliability has been determined. The modification may either increase or decrease the reliability. Second, it is not true after the part has been operating so long that it is beginning to wear out. If this is the case, the reliability will be falling, perhaps quite rapidly. Thus, the procedure is valid only if the parts are operating in that portion of their life in which the failure rate is constant. Next, this procedure is valid only if the test environment exactly duplicates the operational environment. It is often very difficult to simulate the exact operating conditions because one subsystem in a booster frequently affects another subsystem near it in a manner which is difficult to predict. Finally, the procedure relies upon a process of sampling. The parts tested to the point of failure are not the actual parts that will be put in the booster or space system. But it is assumed that because the parts tested are a valid sample from those actually to be used, inferences from the sample to the population from which it was drawn are valid. This may or may not be true. However, in spite of its disadvantages, the procedure of testing the reliability of a system by testing the reliability of the parts is widely used and is generally conceded to yield conservative figures. If one must make errors in estimation, it is better to err on the conservative side.

Improving a Low Reliability

If the reliability of either a system of a component is too low to be acceptable, there are many things which can be done. In fact, most large manufacturing companies have a staff of engineers and statisticians whose sole job is to study the reliability of the company's products and recommend procedures to improve reliability. This requirement within a company is becoming more important because the Government has stated that reliability for a system be one of the design specifications in its contracts.

PRODUCT IMPROVEMENT THROUGH QUALITY CONTROL.—The first group of remedial actions can be called product improvement through quality control. For example, in the process of manufacturing a part, a shop foreman may look at a design and decide that if he were to make a small change, it would be easier to fabricate and would be "just as good." When the change is adopted, the part often turns out to be "just as good" only in the sense that it will work, but its reliability might be seriously degraded as a result of the change. To overcome this natural tendency of workers and foremen to try to improve a product, when they are really not in a position to know whether a change is an improvement or not, many companies forbid any changes at levels of management beneath that of the design engineers.

Another type of quality control action consists of correcting the faulty design of a part which tests at a low reliability. This, of course, should be the work of the design engineer, not of the shop worker. In making the correction, the company must consider the part not only as an entity but also as an integral part of a larger operation.

Sometimes a part functions with high reliability only if it is absolutely clean. Dust, dirt, grease, hair, or other debris, even in the smallest quantity, may cause the part to fail. Examples of such parts are valves, pumps, and lines used to handle liquid oxygen. When foreign particles seriously affect reliability, a company must often use elaborate precautions in the manufacture and handling of the part. The room where work is done may be held under positive atmospheric pressure so that no outside air will seep in, and all incoming air is filtered and conditioned. At the same time, all workers wear gloves and special clothing, and handling is done with special tools. Special cleaning procedures are devised and rigidly enforced. Finally, packaging is carefully controlled and inspected. In fact, the whole process is undertaken in an environment as clean and sterile as that of the hospital operating room.

Often poor reliability of a part or a subsystem is related to lack of employee discipline. It is human to become careless. It is natural to make a few mistakes. In the manufacture of space and missile systems, mistakes and carelessness are too costly. Most companies not only have training programs which impress upon workers the need for good working discipline, but they also have continuing programs which encourage attention to detail and a sense of responsibility.

Quality control is in itself a large subject. It is described here only briefly because it is one of the most important methods for improving the reliability of a part, subsystem, or system.

USE OF REDUNDANCY.—Another method of improving reliability is the use of redundancy. Certain critical parts of a system may be so designed that two or more alternate, or redundant, parts are provided. These parts are so combined that the system fails only if both, all three, or all four of the parts fail. In other words, the parts are combined in such a manner that the system will function if at least one of the redundant parts operates properly.

The calculation of reliability for a redundant portion of a system is exactly the same as the calculation of Example IV under "Contingent Probabilities." Suppose that all the effort of product control has been expended and still a part has a reliability of only .90. An engineer may decide that, to improve the reliability, he will use two such parts, combined in parallel. What is the probability that at least one of these parts will function as planned? Clearly, success of the system is not contingent

upon success of both parts. The probability of failure for each part is $1 - .90 = .10$. The probability that both parts will fail is $(.10)^2 = .01$. The probability that at least one part will operate is $1 - .01 = .99$. Thus by use of a two-path redundant subsystem, the reliability of the part has been improved from .90 to .99.

There are, of course, disadvantages to this plan. First, using two parts instead of one increases weight, and, in space or missile systems, weight is of great importance. Secondly, it increases cost. However, it is a plan that is frequently used because of the advantage indicated. If a two-path redundant system is good, why not use a three- or four-, or five-path redundant system? This is not usually done because the weight and cost penalties increase with redundancy. Further, the gain follows a law of diminishing returns. By using two parts in parallel instead of one part, the reliability was increased by 9% in the example given. But, by using three parts in parallel instead of two, the reliability would be increased by only 0.9%. There may be circumstances under which a gain of 0.9% is enough to compensate for the weight and cost penalties incurred. Thus, the decision to use redundancy and how much redundancy to use is a matter of judgment on the part of the design engineer.

OPERATIONAL TESTS OF RELIABILITY.—So far, reliability has been examined from the point of view of a manufacturer engaged in the process of designing and manufacturing a space system. There is another point of view that must be considered—that of the operational commander who has an inventory of systems and is trying to accomplish his mission with these systems. For example, think of an officer who has the responsibility of a number of ballistic missiles. This officer is keenly interested in the problem of whether or not his ballistic missiles will operate when he gives the order. He needs a good estimate of the reliability of the system he commands.

The commander can obtain an estimate from the contractor, who in turn obtained it by the method of testing component parts just outlined. However, from the point of view of the operational commander, there is another, and perhaps better, method for testing reliability.

Suppose a commander has 100 ballistic missiles in his operational inventory. If he fired them all, he would know the reliability, but he would have no missiles! Obviously, this plan does not solve his problem because he wants to know the reliability of his inventory *before* launch. He has to fire some of the missiles under conditions which closely approximate operating conditions, and then make inferences to what he can expect when the remainder are fired. That is, he samples his inventory, tests his sample, and then makes inferences to the whole population from which the sample was drawn. Clearly, if his inferences are to be valid, his sample must be truly representative of the population. One way of doing this is to select a random sample. He arranges the inventory such that every possible sample has an equal chance of being selected. Suppose an attempt is made to fire a sample of 20 missiles. In the firing, 10 are successes and 10 are failures. The reliability of the sample is 50%. Now, one must make inferences from this sample to the whole population. One cannot simply say that the reliability of the population is 50% because chance might have decreed that the sample had a high proportion of either good or bad missiles in it. In another example, perhaps only 9 would fire. In this event, the reliability of the sample was only 45%. Or, perhaps 11 of them would succeed with the reliability of the sample being 55%. So how can an inference be made from the sample to the whole population?

The way to make inferences has been worked out by statisticians, and their results are summarized in the table below. To check on results from firing 20 missiles, pick out a sample of 20 in the table. Then sclect from the lefthand column the number 10 as the number observed to fire successfully. At the intersection of the column and the row is the number .31. Since the table is built for a 95% confidence level for reliability, this number means that if 10 missiles in the sample of 20 missiles succeed, there would be *95% confidence in the reliability of the population of missiles being greater than 31%*. Another way to express this is to say that, if ten missiles succeed out of a sample of 20, the odds are 19 to 1 that the true reliability is greater than 31%.

Other tables are available for other confidence levels, even as high as 99% confidence. However, the 95% table will continue to be used for illustration. Another way of looking at the situation is that if the sampling process were repeated many times, each time taking a sample of 20 missiles and each time using the table to make inferences from the sample to the population, the result would be right 95% of the time. There would, of course, still be a 5% chance that the inferences would be wrong.

This procedure for estimating the reliability of a system is used by operations officers in many different situations. It has the advantage that one is inferring the reliability from actual testing of the complete system, rather than from testing of component parts. However, it has its difficulties. One difficulty is that it is very hard in practice to obtain a truly random sample. Practical considerations dictate which missiles are fired, and the statistician has to make the best of the results obtained from such practical firings. Second, the missiles fired are not operational missiles fired by operational crews under the stress of battle. Frequently, the missiles used are research and development missiles, and they are fired by specially trained crews supplied by the manufacturer. Inferences from this kind of sample to the operationally fired population is subject to some question. Again, this method of making inferences assumes that the characteristics of the population do not change after the sample is taken. This is seldom true because, if a malfunction is discovered, a fix is attempted.

TABLE 1

95% Confidence Levels for Reliability, p*

	Sample Size, n											
Number of Successes, S	1	2	3	4	5	6	7	8	9	10	15	20
0	0	0	0	0	0	0	0	0	0	0	0	0
1	.05	.02	.02	.01	.01	.01	.01	.01	.01	.01	0	0
2		.22	.14	.10	.08	.06	.05	.05	.04	.04	.02	.02
3			.37	.25	.19	.15	.13	.11	.10	.09	.06	.04
4				.47	.34	.27	.22	.19	.17	.15	.10	.07
5					.55	.42	.35	.30	.25	.22	.14	.10
6						.61	.49	.40	.35	.30	.19	.14
7							.65	.54	.45	.39	.24	.18
8								.69	.58	.49	.30	.22
9									.72	.61	.36	.27
10										.74	.42	.31
11											.49	.35
12											.56	.39
13											.64	.45
14											.72	.49
15											.82	.54
16												.60
17												.66
18												.72
19												.79
20												.86

*Based on Binomial Distribution

One final point should be indicated. Examination of the table will reveal that for the same confidence level, larger samples yield higher reliabilities. Therefore, the larger the sample is, the better; but there is a practical limit to this, too. And there certainly is a practical limit to the smallness of the sample. Clearly, it would be misleading to make inferences from a sample of 1 to a population of missiles.

The discussion just completed has been based upon missiles. Missiles are space systems, at least for part of their operation, and what has been said about missiles applies to all space systems with equal validity. The reliability analysis may be applied to each of the stages of a space system, and the results combined by the laws of probability, or it may be applied to the space system as a whole. The validity of findings on the reliability of space systems takes on new significance when the system is to be manned.

BIBLIOGRAPHY

FRASER, JAMES A. "Reliability," *Air University Review,* XVI, 6 (Sept-Oct, 1965), 23–28.

TAYLOR, GEORGE A. *Managerial and Engineering Economy* (Chapter 14). New York: D. Van Nostrand Company, Inc., 1964.

CHAPTER 11

BIOASTRONAUTICS

THE BROAD scientific field of bioastronautics had its origin many years ago in the daring adventures of men who left the surface of the earth to explore the atmosphere surrounding the earth. Early in these attempts, man learned that flight in the atmosphere presented hazards to life never before encountered. Balloons descended with dead crew members and experimental animals, and, later, the pilots of airplanes were found dead in the wreckage of their craft. There were many speculations, of course, as to the causes of these tragedies, but little was really known other than that high altitude presented a lower barometric pressure and a corresponding reduction of oxygen and nitrogen partial pressures. Obviously, if travel in the atmosphere was to progress, a great deal needed to be learned about the cause and effect of the atmospheric environment. It was at this time that the medical services and physiological and biological researchers launched extensive investigations to determine the characteristics of the environment and the effects of this environment on man. In their research, they relied in part on the work of Claude Bernard, a famous physiologist, who, in 1878, wrote:

> The higher animals really have two environments—an outside environment in which the organism is situated and the inside environment in which the tissue elements live. The living organism does not really exist in the outside world (i.e., in the atmosphere if it breathes) but in the liquid inside world formed by the circulating organic liquid which surrounds and bathes all the tissue elements. The internal environment surrounding the organs, the tissues and their elements, never varies; atmosphere changes cannot penetrate beyond it, and it is therefore true to say that the physical conditions of environment are unchanging in a higher animal. Each one is surrounded by this invariable world which is, as it were, an atmosphere proper to itself in an ever changing world outside. Here we have an organism which has enclosed itself in a kind of hothouse. The perpetual changes of external conditions cannot reach it. It is not subject to them but is free and independent. All the vital mechanisms, however varied they may be, have only one object, that of preserving constant the conditions of life in the internal environment.

What Claude Bernard was saying was that the cell (the unit of living organisms) as found in tissues and organs (muscles, heart, lungs, liver, etc.) of higher animals must remain in a relatively stable environment, regardless of the external environmental changes surrounding the organism. We know that the external environment lies outside man and includes such influences as air pressure, humidity, temperature, odors, vibrations, noise, and acceleration. The internal environment consists of the conditions within the body that affect the activities of cells, tissues, and organs. These internal conditions include such influences as body temperature, blood oxygen, carbon dioxide in the blood, and blood chemistry in

general. Human beings can withstand moderate changes in the external environment without any deterioration in performance primarily because their internal environments are kept reasonably constant by regulatory mechanisms controlled by the central nervous system. These mechanisms provide for exchanges with the external environment to maintain body temperature, blood chemistry, etc., within rather narrow limits. This whole process of maintaining a relatively constant internal environment is called homeostasis.

Changes in the external environment, however, can prove too extreme for the regulatory mechanisms. In such cases, these mechanisms are unable to maintain a constant internal environment, and there results a deterioration in the performance of the sense organs, central nervous system, and/or muscles and glands. Of course, when the human being is a part of the man-machine system, the performance of the system deteriorates as well. This is essentially what the early aviation physiologists and aviation medical personnel had to work with in the infancy of their endeavor. Since their beginning they have amassed a large volume of information designing or outlining all of the known parameters of the external environment and their effects on man's internal environment and his regulatory mechanisms. Definite tolerance limits to external forces have been detailed for physiological functions in the human.

Therefore, we now have known values or quantities of physiological functions which can be applied to a known value or quantity of external environmental stress. If the value of the physical stressor exceeds the ability of the body to adjust to it, we must either reduce the stress or provide some support to the regulatory mechanisms of man in order for him to tolerate the stress and remain homeostatic.

Now that we have entered the new space age and man is attempting flight into the hostile environment of space, we again look to the physiological research field for help in meeting the stresses placed on man by travel into this new environment. This field of study is called bioastronautics, which is the study of the physiological and psychological problems facing man in space travel.

The problems are many but, thanks to the diligent research of physiologists and medical personnel, some of the answers to these problems in space flight were available long before man's first orbital flight. These answers were based on the data gathered in the preceding fifty years of aviation medicine and physiological research. Thus, although the name bioastronautics is new, the techniques and procedures used to determine man's tolerance to external stress are well known, and space flight by man is relatively predictive and safe as a result. To understand these parameters more fully, the major physiological stresses placed on man during space flight must be scrutinized.

PHYSIOLOGICAL STRESSES

As mentioned before, changes in the external environment may at times be very severe and too extreme for man's regulatory mechanisms. These changes cause stress on these mechanisms which is nothing more than pushing the homeostatic levels of man to the tolerance limits or beyond. As man ventures into space, he is exposed to a series of stresses, each of which could prohibit manned space

flight if no corrective measures are available. This chapter discusses, in some detail, most of the major environmental stress parameters, how they affect man, and how such forces are overcome so that man can maintain his bodily functions within his homeostatic tolerance limits. In general, there are two types of environment in which changes can cause severe stress—the physical environment and the mechanical environment. Stresses in the physical environment are caused by changes in atmospheric conditions, and stresses in the mechanical environment are caused by operation of the space vehicle. Each of these environments is discussed separately.

Physical Environment

The earth is surrounded by an atmosphere approximately 100 miles thick. This atmosphere provides ambient pressures, temperatures, humidity, and certain gases required in man's normal environment. It also provides a shield which prevents certain matter and radiant energy from reaching the earth's surface.

As man progresses from the surface of the earth upward through the sea of air, the atmospheric pressure, gas concentrations, and humidity decrease and the temperature varies widely. Above the atmosphere, man is no longer protected from meteorites and various forms of radiant energy that are found in space. These changes in the external physical environment are too extreme for man's regulatory mechanisms and he cannot exist without a variety of protective devices.

For protection from a lack of oxygen and pressure, the vehicle which transports man into space is a sealed cabin which provides 5.5 psi environmental pressure of 100 percent oxygen. There are several advantages in using a lowered pressure of pure oxygen. First, it will provide the lungs with sufficient oxygen partial pressure to assure saturation of the red blood cells (RBC's) which transport oxygen to all living cells in the body. Secondly, the nitrogen-free atmosphere helps "wash" nitrogen from the tissue fluids. The elimination of the nitrogen that is dissolved in the tissue fluids is important because, if the body is exposed to an ambient pressure equivalent to the pressure at 35,000 ft altitude, the dissolved nitrogen can be released as a gas. If this released nitrogen becomes trapped in the joints, it causes severe pain and stress called disbarism. The third advantage of using the low pressure, pure oxygen system is that there is a large savings in the structural weight of the vehicle because it must be stressed only to withstand a 5.5 psi pressure differential rather than 14.7 psi.

HEAT AND HUMIDITY—Man's internal heat regulatory mechanisms tend to keep his body at a constant temperature of about 98.6° F. Normally, the body loses heat constantly through the lungs and skin. But, if a person needs to lose more heat to maintain constant body temperature, his regulatory mechanisms come into play to increase the heat loss. These mechanisms include dilation of the blood vessels near the skin (bringing more blood near the surface of the skin) and perspiration which increases heat loss due to evaporation. Conversely, if a person's environment cools and he needs to conserve heat, perspiration stops and the blood vessels constrict. In addition, if the cold becomes extreme, reflex shivering causes the muscles to produce more heat. Normally, thermal balance can be easily maintained by a nude body if the environmental temperature is

70-80° F and the relative humidity is about 45 per cent. But, in spacecraft operation, many factors come into play. The flux of heat energy is large, and the regulatory mechanisms employed by man to maintain his body temperature are inhibited. The sources of heat energy, other than from man himself, include the electronic equipment in the vehicle, friction heat as the vehicle leaves or reenters the atmosphere, and heat energy from the sun which is no longer dissipated by an atmosphere before reaching the vehicle. Because the heat sources produce a high heat load and the man's normal heat loss mechanisms are reduced, additional protection is required. Special suits are worn under the pressure suit to provide circulating air or water to help dissipate the heat, and air conditioners provide additional protection by regulating the ambient cabin temperature.

CONTAMINANTS.—Contaminants, such as carbon dioxide, methane, and others, become a problem in space capsules due to the confined space and the sealed characteristics of the vehicles. All atmospheric substances must be considered to be toxic if introduced into the body in amounts greater than some threshold value. Each contaminant, therefore, must be treated individually in terms of its own threshold value. In all cases, both concentration and the time of exposure are critical conditions. It is beyond the scope of this chapter to cover the multitude of contaminants that are possible in a space capsule. Therefore, it is sufficient to say that all substances normal to a space cabin environment must be kept at non-toxic levels even to the point of removing some substances entirely. More will be said of certain substances later in this chapter.

RADIATION.—The types of radiation which man will encounter in space were discussed in the chapter on "Space Environment." This chapter discusses the hazards to man from the wide range of radiant energy to which man will be exposed when he leaves his protective blanket of the earth's atmosphere. The ultraviolet, visible, and infrared radiations found in space could cause problems for the crewmembers if they were unprotected. However, the spacecraft, their suits, and protective visors have virtually eliminated all hazards from these radiations. The main hazards to life are found in the ionizing radiations and their effect on living cells. The damage to a living cell can be explained through a consideration of the physical changes that occur within the cell. It is known that the normal structure and the electrochemical makeup of biological material can be altered by radiant energy. This results in the breakdown of the molecular structure and/or the production of toxic substances. These changes usually are lethal to cells, especially if the enzymes, which are catalysts for all biological reactions, are destroyed. If the cell chemistry is disrupted, the cell dies. If sufficient cells of an organ are damaged, the organ will cease to function and the man will die. The biological effects of radiation can be summarized as follows:

1. Manner. The effects may be direct (direct damage to the nucleus of a cell) or indirect (damage to the enzymes and chemistry of the cell).
2. Time. The effects may be immediate or delayed depending on the *manner* of damage.
3. Effectiveness. The effects may be reversible or irreversible depending on the amount of radiation and the kind of tissue involved.
4. Systems. There are three main organ-systems sensitive to radiation: the blood system, digestive system, and the central nervous system.

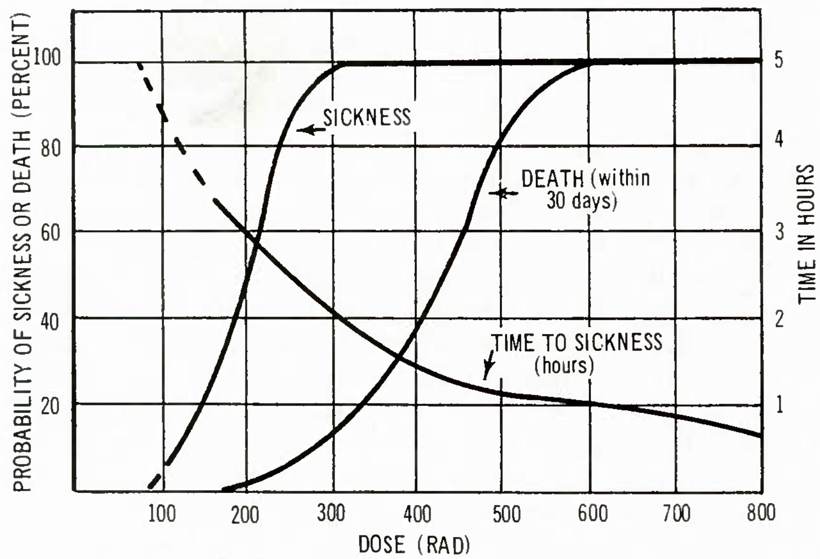


Figure 1

One unit of measurement used in relating radiation effects in biological materials is the Radiation Absorbed Dose (RAD). One RAD is equivalent to the absorption of 100 ergs of energy per gram of living tissue. The known effects of acute whole-body radiation in RAD are shown in Figure 1.

As indicated, if a population receives approximately 225 RAD of acute, whole-body radiation up to 50 percent of those individuals will become ill within three hours. However, the tolerance differences between individuals is quite wide, and

DOSE ACCUMULATION RAD

ZONE	8 – 10 gm/cm² SHIELDING	TRAINING 1 – 3 DAYS	EXPLORATION 1 – 2 WKS	2 – 4 WKS	APPLICATION 1 – 3 MONTHS
1	LOW EARTH ORBIT	0.06 RAD	0.28 RAD	0.56 RAD	1.8 RAD
2	ANOMALY	0.5 "	7.0 "	14.0 "	45.0 "
3	2 PENETRATIONS OF NATURAL BELTS	10.0 "	10.0 "	10.0 "	10.0 "
5	GALACTIC BACKGROUND	0.045"	0.2 "	1.0 "	3.0 "
	TOTAL	~11.0 "	~18.0 "	~26.0 "	~60.0 "
4	SOLAR FLARE	100 RAD	100 RAD	100 RAD	100 RAD

Figure 2

the rate of onset of illness can vary considerably. If the dose received is approximately 450 RAD, up to 50 percent of the population will die and the other half certainly will become ill. But what kinds of doses can we expect in normal space operations: Figure 2 gives the expected dose levels to man in a space vehicle with 8-10 grams/cm^2 shielding.

Two factors are apparent. The dose levels are rather small and the rate of onset of the dose levels are spread out over days, weeks, and months. Based on this information, Col J. E. Pickering, USAF, MC, has proposed mission selection according to exposure rates. His proposal considers the age of the crewman, the number of years which he expects to spend in the flight program, and the provision of adequate recovery periods after each exposure. He assumed that space programs would continue to use a space-qualified crewman for as long as safety of the crewman can be assured. In this approach, an older man (35 years and older) would be permitted only one year of flight. The schedule would call for one week of rccovery for each RAD received by the crewman. In the case of an older crewman, the radiation exposure would need to be very carefully managed to gain maximum usage of his flight time through training, exploration (space missions), and applied (routine repeated flights) phases. A high rate of exposure during training or exploration would deny his experience for routine applied missions. In the case of younger men who will have three to five years of flight available, proper management of their radiation exposure will provide a long useful career in the exploration and applied phases of operation. Pickering proposed that a total dose of 80 RAD distributed over the five-year career of the younger man could be tolerated as long as a compensatory one-week recovery time for each RAD received was provided after each period of exposure. This suggested dosage still permits an additional emergency dose of 100 RAD due to a solar flare without serious jeopardy to mission success. This seemingly large total dosage is considered reasonable because the numbers of the population exposed over the next several years will be small and carefully observed. Thus, the added risk due to radiation above and beyond the total risks associated with the mission will remain extremely small. Pickering's reasoning behind the approach is to return flexibility to the operational mission as far as radiation is concerned (see Figures 3, 4, 5).

Mechanical Environment

The problems associated with the mechanical environment includes all those stresses placed on man as a result of the vehicle and its operation in the space environment.

Acceleration.—As the rocket engines thrust the space vehicle toward orbital velocity and as the vehicle changes velocity during reentry, a significant increase in acceleration forces is experienced. These forces are measured in units of gravity forces (g) with one g being equivalent to the force of gravity acting on a body at sea level. Booster accelerations are unique for each stage and for each type of vehicle. The peaks of acceleration enroute to orbit range from 3 to 8 g's. The reentry acceleration normally does not exceed about 8 g's. The acceleration forces for a typical manned vehicle flight profile are shown in Figure 6.

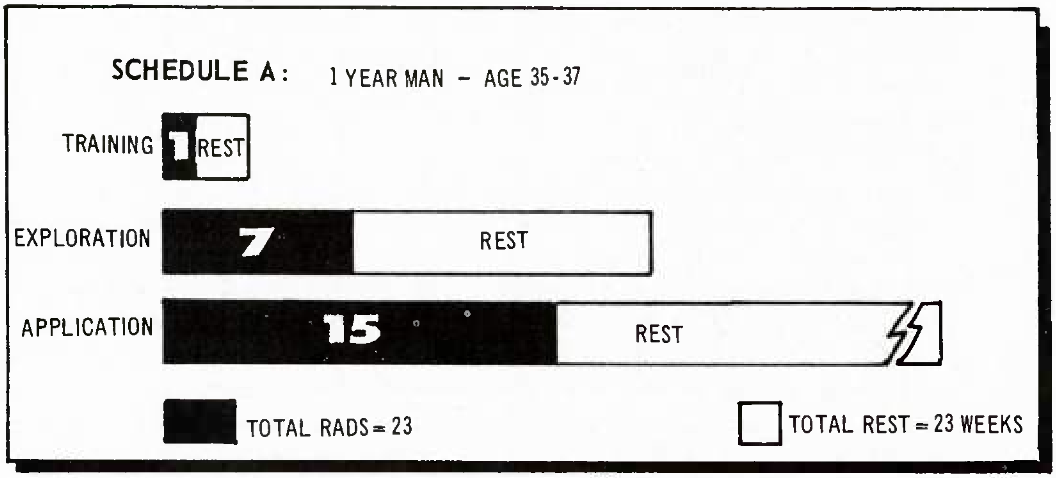


Figure 3

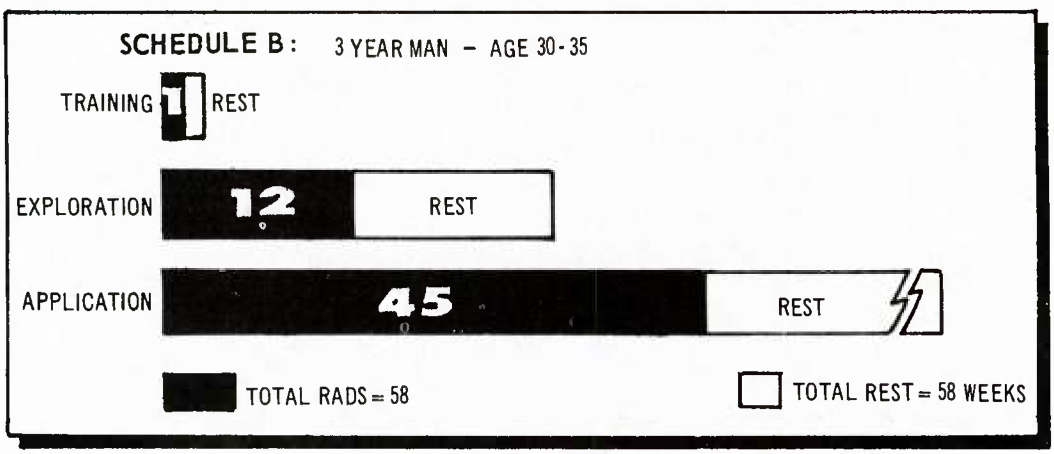


Figure 4

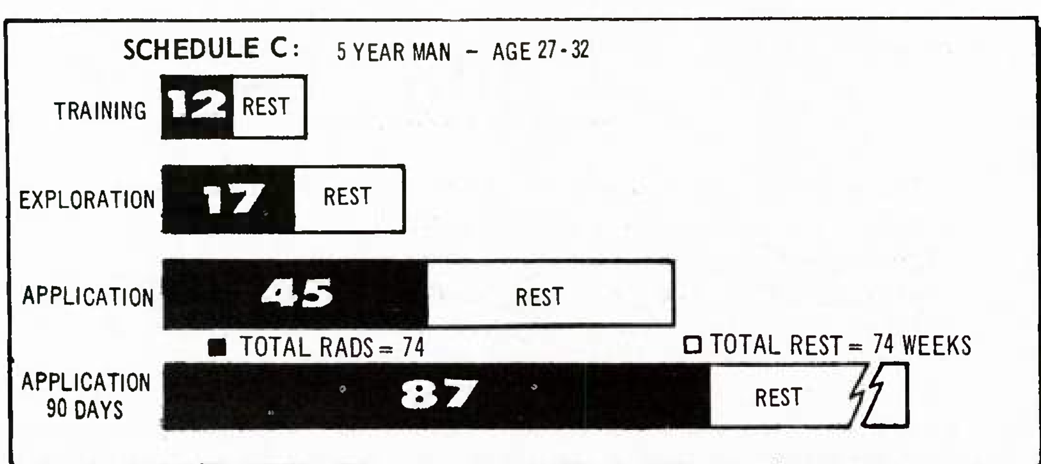


Figure 5

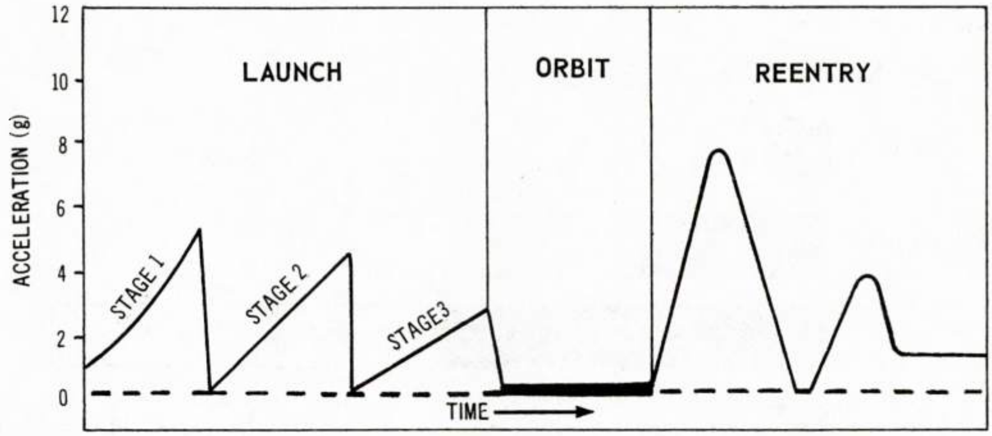


Figure 6

Because of the effects of increased forces on the cardiovascular and the muscular system, man must have some protection. The circulatory system tolerance varies, depending on the manner in which the g force is applied. When the acceleration is positive (from head to foot), an average maximum tolerance of 5 g's for very short periods of time or 2 to 3 g's over longer periods can be expected. Negative g forces (foot to head) are less tolerable and a maximum of 2 to 3 g's is the tolerance level. Each of these tolerances would be marginal for space flight operations. Positive and negative g forces cause pooling of the blood in the extremities because the heart is unable to overcome the acceleration forces and complete the circulatory process. However, these conditions can be avoided by properly orienting the individual in the vehicle so that the g forces are transverse to the axis of the body—applied from front to back or back to front. In this position, the g force is not acting on the long hydrostatic columns of blood that

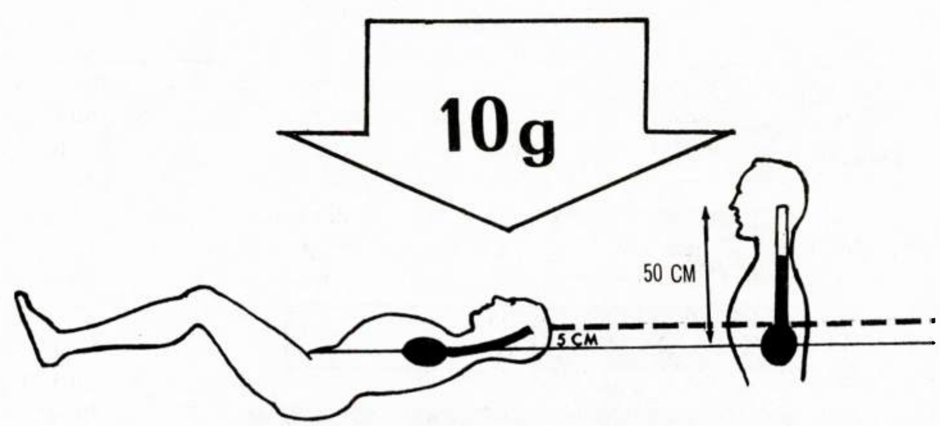


Figure 7. In the spacecraft, acceleration protection for the cardiovascular system is given by reclining in a couch so that the force is transverse. At 10g blood can reach the head with sufficient pressure to perfuse the brain. In the upright position, the blood will not reach the head.

exert heavy pressures on the heart. In fact, even at high transverse accelerations, the peak g load on the heart should, with proper positioning, be only about 1 to 2 g's (Fig. 7).

Therefore, our present manned systems place the crewmember in a reclining position, with the legs and head raised slightly. With the acceleration forces acting in the transverse direction, the man can withstand about 20 g's for short periods of time.

The arms, legs, and head also present difficult problems under high acceleration forces. For example, the arms cannot be controlled efficiently if g forces exceed about 4 g's. The hand and fingers have tolerance levels of about 7 to 9 g's. The astronaut's control and operation of the spacecraft would be limited or impossible if he used conventional aircraft controls. Therefore, new controls and control locations were devised so that the astronaut would have no difficulty operating his vehicle under high acceleration forces. These include side-stick controls for pitch, roll, and yaw of the space vehicle.

VIBRATION.—A man in a space vehicle during launch and the rapid flight through successive layers of atmosphere can be exposed to rather severe vibrations from the rocket motors and buffeting in the atmosphere. Vibrations are side to side jostling motions and up and down bouncing. The man is most sensitive to vibration frequencies of 1 to 10 cycles per second because the major internal organs have a natural resonance frequency in this range. When the organs become resonant with the vibrations of the space vehicle, severe pain, nausea, headaches, and dizziness are common symptoms. The organs begin to tear away from the mesentery holding them in place. Therefore, the vibrations in the frequency range of 1-10 cycles per second must be eliminated or dampened in order to make space flight safe for man. Under present power and flight schedules used in manned flight, vibration problems are sufficiently controlled and produce no serious drawbacks to the program.

NOISE.—Noise has been a problem to man for a long time, but never before has the intensity been so great as the noise produced by space boosters. It has been known that men working in noisy environments develop a restricted deafness in the frequency range of the noise. It is necessary to place very rigorous protective controls on workers, including the use of ear plugs and ear muffs. Man's maximum tolerance to noise is about 140-150 decibels at which level permanent damage to the hearing mechanism will occur if exposed for only about one minute. Space boosters produce noise levels of 145 to 175 decibels. The solution to the noise problem involves using some of the natural physical conditions of space vehicles. For example, the capsule is placed on the end of the booster away from the rocket motors and, since distance attenuates noise, the decibel level in the capsule is reduced. Also, the materials in the vehicle, including the skin, structure, propellants, and oxidizers, as well as the equipment on board the spacecraft, help attenuate the noise intensity. Finally, the space suit, which is worn during the operation of the first stage booster, has sound and vibration reduction materials throughout the suit and in the helmet. These techniques have been employed successfully in the manned space programs, and the noise level in all vehicles has been kept well within the limits of safe operation.

"WEIGHTLESSNESS."—There were many anxieties concerning the effects of "weightlessness" on man operating in space. These ranged from fear of falling, nausea, and injury to the more complex physiological considerations, such as disruption of normal biological functions. Most of these anxieties have been dispelled as our knowledge and experience have increased. Although much has been learned about the effects of "weightlessness" on man, we still must proceed cautiously as we expand our capability. At the present time, we can divide the problems posed by "weightlessness" into two categories: psychological and physiological.

The psychological problems are many and varied. Factors such as isolation, restraint, and confinement apparently have presented no particular problems in our shorter flights of up to 14 days duration. However, as we anticipate longer flights, these three factors may present some serious problems. This was emphasized by Col Borman on the Apollo 8 lunar orbiting mission, when he said, "We found several things we think need to be improved. I think that we need to concern ourselves with proper engineering for better body waste disposal systems. We have to provide showers on board. We have to have some sort of entertainment, television or canned tapes. We have to look into better foods. We have to realize that when you put a man on orbit for 60 days, or perhaps even longer, you have to pay more attention than we have in the past to his basic *creature comforts."*

Obviously, man's psychological weaknesses are becoming important, even on relatively short-term flights. He wants basic "creature comforts" to be examined closely. The isolation, restraint, and confinement factors of a space vehicle are beginning to take their toll. Similar reactions generally occur when test subjects are confined from 10 to 14 days in a small room with limited comfort facilities and with only a work-rest schedule.

Several projects are underway now to develop equipment designed to overcome the deficiencies in "creature comforts." Recorded tapes and even space oriented games are being studied to provide entertainment. Showers, consisting of plastic bags and pressure water sprays, are being tested. More efficient body waste disposal systems are being developed, as are new methods of preparing appetizing foods.

Another area of concern is the sleep pattern for space flight. No doubt the normal "physiological clocks" of the astronauts are upset during the flight schedules experienced so far. For example, in low earth orbit, the astronauts experience approximately 15 day-night cycles in a 24-hour period. Also, when in "weightless" flight, the astronauts seem to need less sleep than they do in the 1-g environment of earth. As a result, there has been considerable juggling of the work-rest schedules in our missions to find the optimum regime. But, because all missions vary so much, finding an optimum schedule has not been an easy task, and no clearcut timetables can be made.

The psychological stresses facing man in space flight have not been too great thus far. In the future, when flights of months and years become a reality, we may face serious psychological barriers. Much study is necessary to insure safe space flights for long periods of time.

There are many physiological and adaptive changes that occur when man is exposed to a zero-g environment. In fact, this subject is one of the most controversial and possibly the least understood area related to manned spaceflight. Some feel that the problems involved will be the crux of the question concerning man's ability to operate in the space environment. Others feel that man's adaptive processes are so great that there will be no problem in placing man in the "weightless" environment for extremely long periods of time.

It is known that as man lives in the presence of the earth's gravitational field it produces certain effects within his body. There is relative displacement of organs, tissues, and fluids as a result of their different densities and varying orientations to the direction of the gravitational force. If there are any physiological changes resulting from "weightlessness," they must be the result of the apparent removal of the gravitational force. The question is: what are the effects? Gravity is so strongly associated with the normal human physiological activities that the total effects of its removal are difficult to determine. We have experienced several thousand hours of "weightless" flight and have noted changes in several of the organs and organ systems, but they have not made the environment intolerable.

Cardiovascular deconditioning is probably the most significant problem. We have noted changes in blood volume, blood fluid shift, changes in the cellular ratios, and general deconditioning of the muscles and elastic properties of the blood vessels. These changes occur in the first 10 to 12 hours of exposure and are directly related to the "weightless" state of the blood and, as a result, the general reduction in work required to move blood throughout the body. There seems to be no particular problem adapting to the "weightless" environment, and it is possible that the system tends to reach a new steady state, at least in the short flights to date. The cardiovascular changes that occur also seem to be within the tolerance limits of a normal individual for the periods involved. The major problem arises when the astronauts reenter the earth's atmosphere. They will experience high deceleration forces and rapid return to the normal force of gravity on the cardiovascular system. It is possible they may experience orthostatic intolerance and other related stresses until the cardiovascular system can readapt to the gravity environment.

Muscle deconditioning and body mineral imbalance are other physiological disturbances that have appeared during "weightlessness." There has also been some discussion of possible psychophysiological disturbances related to long term "weightlessness" such as degradation of alertness and attention, vestibular function, and long term loss of vestibular and proprioceptive sensory information. Although not seriously upsetting at this time, it is possible that these functions in long term flights when combined with other psychological and physiological stresses may cause serious problems.

In search for methods of overcoming the debilitating effects of "weightlessness," it was found that the symptoms were remarkably similar to those of immobile bed patients. These patients suffered blood volume changes, fluid shifts, orthostatic intolerance, weakened muscles, and calcium loss. The treatment for these ailments has been exercise of the patient within his capability. Similarly, regular and vigorous preconditioning of astronauts and regular in-flight physical exercise

have prevented the physiological changes from going beyond the tolerance limits. However, on future longer flights it may become necessary to provide artificial gravity for astronauts by slowly rotating large space stations.

LIFE SUPPORT EQUIPMENT

Without question, space presents an extremely hostile environment for man. In order to survive and operate a space vehicle, an astronaut must have adequate protective devices to allow him to remain in his homeostatic condition. Life support equipment designed to give him protection during space flight includes biomedical, ecological and environmental items.

Biomedical

In short or long term space exploration, we must be able to monitor the health of an astronaut and prescribe medication if the need arises. As evidenced on recent Apollo missions, the transfer of germs from one astronaut to another in the close confines of our present space vehicles is rapid and unavoidable. Therefore, medical monitoring equipment is essential. Modern medical instrumentation allows monitoring of many physiological functions on a real time basis. For the purpose of immediate health monitoring, a biomedical belt or harness has been designed. This harness can, with the proper electrodes and receiver-transmitter units, relay the data on a variety of physiological functions to the earth sensing stations via telemetering methods. These functions include the electrocardiogram, blood pressure, respiratory rate, electroencephalogram, and body temperature. Under the present program, only heart and respiratory rates are monitored in real time. Other medical tests are performed before a flight begins and when the crew returns to the earth. These tests analyze urine, fecal, and blood samples. The future will possibly bring increased monitoring of the crewmembers on long term flights. This will be true especially when flights begin to carry men to the distant reaches of our planetary system wherein there will be time delays in transmitting information back and forth between the space vehicle and the earth.

Should it be necessary to administer medication, the astronauts have rather complete medical kits on board. These small and compact kits (6 by 4.5 by 4 in) contain motion sickness and pain suppression injectors, first aid ointment, eye drops, nasal sprays, compress bandages, adhesive bandages, thermometers, and a variety of antibiotic, nausea, stimulant, pain, decongestant, diarrhea, aspirin, and sleeping pills. On the advice of medical personnel, the crewmembers can use the medicines in the kit to relieve most problems anticipated on present missions. Medical kits can be varied as the mission varies. However, if a medical problem is beyond the capability of the kit and does require medical attention, the mission will be terminated as soon as possible.

Ecology

Ecology by definition is the study of the relationship of organisms to their environment. This means the study of all cyclic environmental factors related to

the normal life cycle of living organisms, such as atmosphere, pressure, temperature, water, light, food, life expectancies, predators, waste disposal systems, etc. As one can imagine, the relationship of man to his environment is very complicated and becomes an even greater problem when an attempt is made to pack his normal ecology into a small space vehicle. A very much simplified description of one cyclic pattern can be shown with the chemical equations for photosynthesis and respiration. Green plants, those containing the substance chlorophyl, can, in the presence of sunlight, combine carbon dioxide (CO_2) and water (H_2O) to make glucose ($C_6H_{12}O_6$) and oxygen (O_2). Animals can metabolize glucose and oxygen to produce carbon dioxide and water. The cycle is complete.

$$
\begin{array}{lcccccc}
\text{Plant photosynthesis:} & 6CO_2 & + & 6H_2O & \xrightarrow[\text{Chlorophyl}]{\text{Sunlight and}} & C_6H_{12}O_6 & + \; 6O_2 \\
 & \uparrow & & \uparrow & & \downarrow & \;\;\downarrow \\
\text{Animal Respiration:} & 6CO_2 & + & 6H_2O & \xleftarrow{\text{Metabolism}} & C_6H_{12}O_6 & + \; 6O_2
\end{array}
$$

The balance of the many cycles which make up our closed ecological system on earth is, of course, much more complicated than shown with the sample equations.

The approach to life support equipment design is based on the analysis of human requirements on the one hand and the synthesis of the equipment to provide protection within the human tolerances on the other. To learn how the environmental control systems are developed, let us first discuss the food, oxygen, and water requirements for man.

As shown in Figure 8, in an open ecological system, man has a requirement for approximately 22.7 lbs of oxygen, 4.7 lbs of water, and about 1.3 lb of food per day. At the same time, he produces an equivalent weight of waste products, including 20.7 lbs of oxygen, 5.2 lbs of water, 2.2 lbs of carbon dioxide.

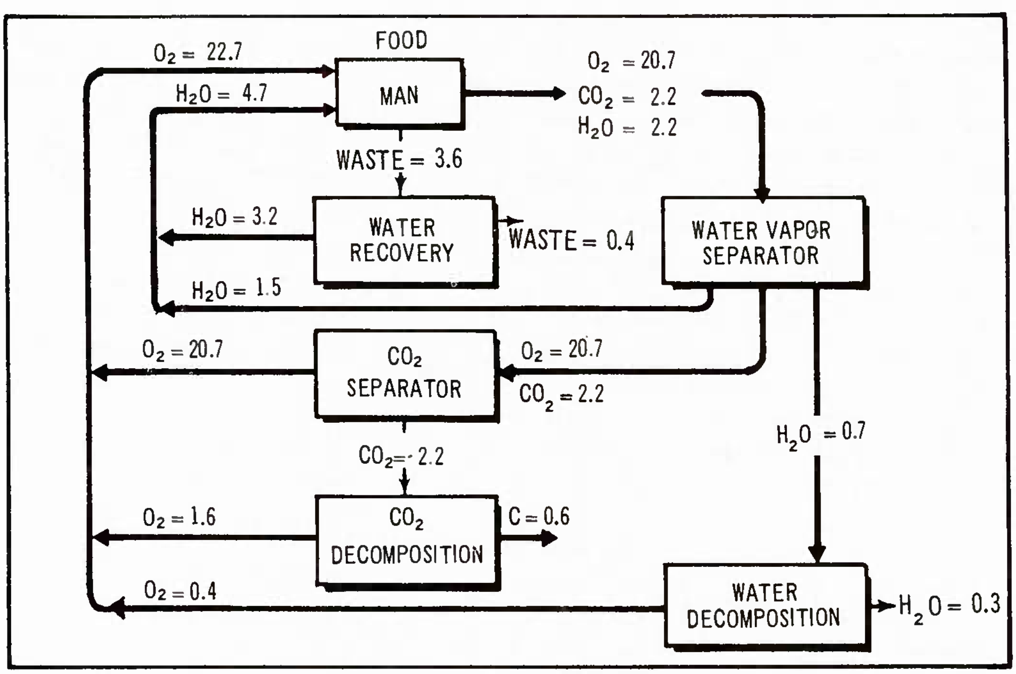
OXYGEN – WATER – FOOD EXCHANGE		
INPUT LBS/DAY		OUTPUT LBS/DAY
OXYGEN (O_2)	22.7	20.7 (O_2)
WATER (H_2O)	4.7	5.2 (H_2O)
		2.2 WATER VAPOR
		3.0 URINE
FOOD	1.3	2.2 CARBON DIOXIDE (CO_2)
		0.6 SOLID WASTE
	28.7	28.7

Figure 8

At the present time, it is impossible to carry these stored materials on space missions because of the weight involved. Some progress has been made in weight-saving in the food program. Because of the mechanical problems of eating food during weightlessness in orbit, special preparation and packaging is necessary.

Our present food system consists of dehydrated foods prepared by a process of freeze-drying. This water-free food is then packaged in special plastic bags. The food can be rehydrated from the hot or cold water on board the spacecraft or, in the case of some items, rehydrated by saliva in the mouth. This current food system has the advantages of being bacteria-free, lightweight, easily stored, and easily prepared. When rehydrated, the food is tasty and nutritious. However, it still has the disadvantage, inherent in an open ecological system, of being impractical from the viewpoint of long flights of months and years. Therefore, some system of regenerating or growing food from the materials on board the spacecraft must be designed for future long term missions.

The oxygen required by man also requires investigation. In an open ecological system, oxygen is man's largest waste product. He breathes in 22.7 lbs of oxygen per day, but, as shown in Figure 8, he uses only 9 percent of the oxygen inhaled and throws away 20.7 lbs of the oxygen. This is the largest portion of the total waste output by man. Therefore, to make spaceflight economically feasible from a weight of storables standpoint, better treatment of the gas portion of man's system is required. To do this, the exhaled waste gases are processed to remove the carbon dioxide and water vapor. The remaining oxygen is then returned to the astronaut.



CLOSED GAS AND WATER SYSTEM

Figure 9

Research physiologists and biologists, therefore, are working to close all the various loops of an open ecological system. Figure 9 shows closed loops for oxygen and water.

This system does not account for the astronaut's food. However, food can be produced by using algae or bacteria in a system where these organisms will convert the CO_2 and H_2O wastes into foods high in protein and carbohydrates. These organisms will also use (as they do on earth) the urine and fecal material of man in their nutrient broth. This system will provide all the oxygen and food requirements for man and rid him of his waste products. The required water would come from the fuel cells of the spacecraft. Eventually, a completely closed and balanced ecological system will be designed which will include man as an essential loop. His waste products will be the food for the microorganisms, and their waste products will be the food, water, and gas requirements of man. This is the system under which we are now living on earth.

Space Suits

In addition to the cabin environmental control systems, environmental suits must be provided to protect the astronauts in the event of loss of cabin pressure and also for extravehicular activity (EVA). Because the environmental factors do not change for man, the suit must provide the same protection as was provided by his cabin. Therefore, an undergarment has been developed that provides spacing for a cooling flow of oxygen around the body and attachments for a communication belt, urine collection equipment, and the biomedical instrumentation belt. This suit is quite comfortable and worn during all activities in the space vehicle. Before the astronaut begins EVA, he switches to a similar underwear type suit in which small tubes are placed throughout the suit to carry cool water which dissipates the body heat. Over the undergarment, the astronaut wears a combination suit for pressure, micrometeorite, and radiation protection. The pressure portion of the suit provides 5.5 psi pressure to all parts of the body except the head. Here oxygen pressure is provided at 3.75 psi by the helmet. The helmet, in addition to the communication equipment, provides filters of various densities to protect the astronaut from the blinding glare of the sun. The multi-layered external portion of the pressure suit provides protection against radiations and micrometeorites. This material is made of two inner layers of Beta cloth to intercept micrometeorites, three to four layers of insulating cloth, a layer of special metalized fabric, and another layer of Beta cloth. The material and design of the suit does not inhibit movement severely but does provide sufficient protection for activities outside the space vehicle.

BIBLIOGRAPHY

Aviation Week and Space Technology, LXXV, 7, Aug 15, 1966.

GERATHEWOHL, S. J., *Principles of Bioastronautics.* New York: Prentice-Hall, 1963.

LANGLEY, L. L., *Outline of Physiology.* New York: McGraw-Hill, 1961.

MORGAN, et. al., *Human Engineering Guide to Equipment Design.* New York: McGraw-Hill, 1963.

NASA Contractor Report, NASA CR–1205, *Compendium of Human Responses to the Aerospace Environment,* Nov 1968.

Appendix A

MATHEMATICS REVIEW

IN DESCRIBING the orbit of an artificial satellite, engineers and scientists sometimes find it convenient to use Cartesian coordinates. In following this system, they locate the satellite in terms of three intersecting lines, or axes, which are referred to as the x-, y-, and z-axes. The origin, or point at which the three axes intersect, is usually placed at the center of the earth. One axis is the earth's rotational axis. The other two axes would lie in the earth's equatorial plane. These Cartesian coordinates are then related to another set of coordinates used to locate the sun, the planets, and the other heavenly bodies.

Another way of locating an artificial satellite is to specify its orbital elements. Orbital elements are discussed in Chapter 2.

Before locating artificial satellites or ballistic missiles in terms of three dimensions or of motion, we might do well to simplify the problem by considering the location of any two points in the same plane on the surface of the earth. These points might be designated through the use of either rectangular or polar coordinates.

Rectangular Coordinates

A rectangular coordinate system is established by first drawing two perpendicular intersecting lines, or axes. It is natural to locate position in this way because we think of the two axes as directions defined by the points of the compass. In giving directions in the rectangular coordinate system, we might say, for example, "To get to the post office from your present position, go five blocks south and two blocks east." The present position is the origin, and dirction plus distance identifies the position of the post office (Fig. 1).

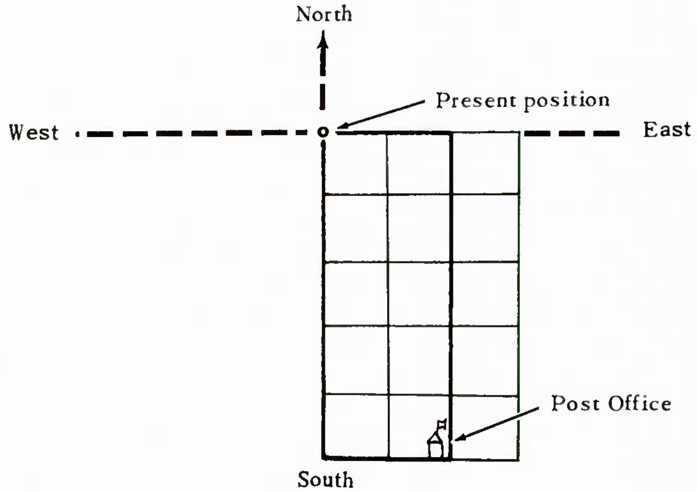


Figure 1. Rectangular grid system.

In a rectangular coordinate system the horizontal line is called the x-axis, or axis of abscissas, and the vertical line is called the y-axis, or axis of ordinates. The origin, or center, of the system is the point of intersection of the axes and is usually labeled O. Coordinates of a point are measured parallel to the axes, using the origin as a starting point. Positive x values are measured to the right of the origin and negative to the left; positive y values are measured upward from the origin and negative downward. The arrows on the axes indicate the positive direction. The axes divide the plane of the coordinate system into four quadrants numbered counterclockwise. Quadrant I is the upper right of the system.

Points are identified in the system by assigning an x value first, then a y value; for example, P_1 (4, 3). The Point P_1 (4, 3) is 4 units to the right of the y axis and 3 units above the x-axis. The Point P_1 and several others are plotted on Figure 2. A point in general is denoted as P (x, y).

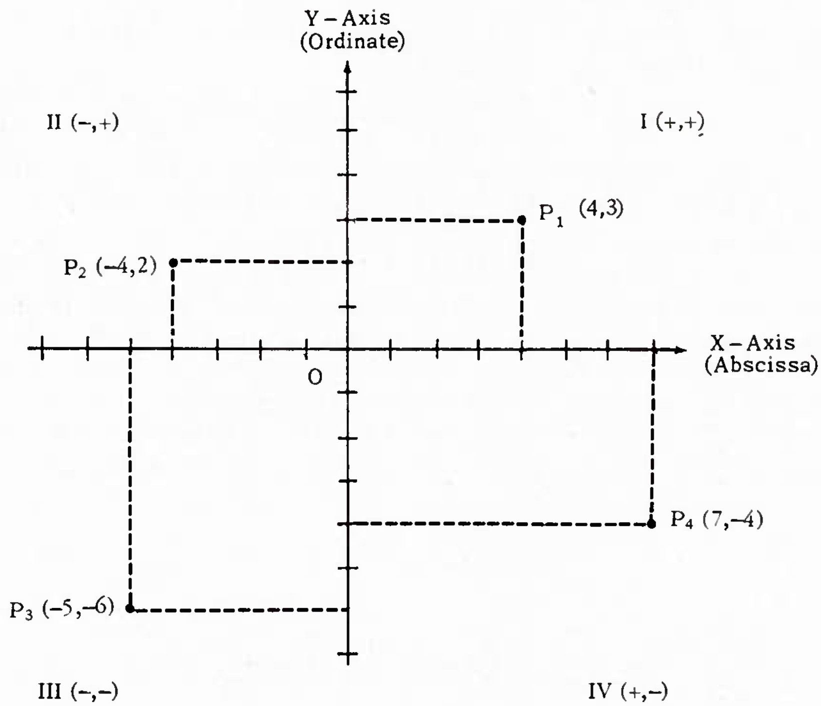


Figure 2. Rectangular coordinates.

Vectors, which are directed line segments, are easily adapted to rectangular coordinates. For complete description, many physical quantities, such as force, velocity, acceleration, and displacement, require a specification of both magnitude and direction. These quantities are known as vector quantities and must be treated mathematically as vectors. When such quantities are represented as vectors, the length of the line is proportional to the magnitude, and the direction is the direction of the vector quantity.

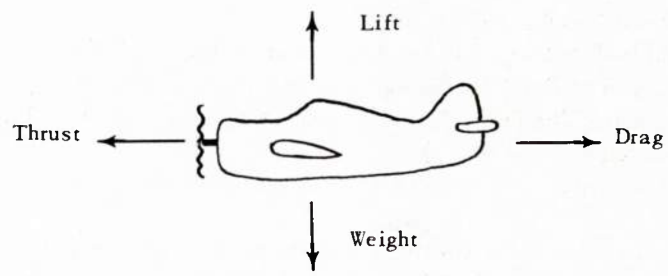


Figure 3. Forces acting on an aircraft in flight.

Vector addition can be shown graphically by drawing successive vectors each with its tail on the head of the preceding vector, and the resultant is drawn from the tail of the first vector to the head of the last vector. For example, if vector v is to be added to vector u, the resultant w is drawn as shown in Figure 4. Note that the resultant w is a vector generally acting in a new direction. Its magnitude does not equal the magnitude of vector u plus the magnitude of vector v.

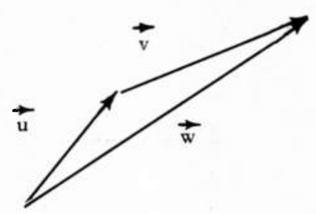


Figure 4. Vector addition shown graphically.

Fortunately, vectors do not have to be handled graphically but can be separated into components which are simple to handle in rectangular coordinates. Each vector is separated into two components, one parallel to the x-axis and one parallel to the y-axis. Thus, in Figure 5

$$u_x + v_x = w_x, \text{ and } u_y + v_y = w_y$$

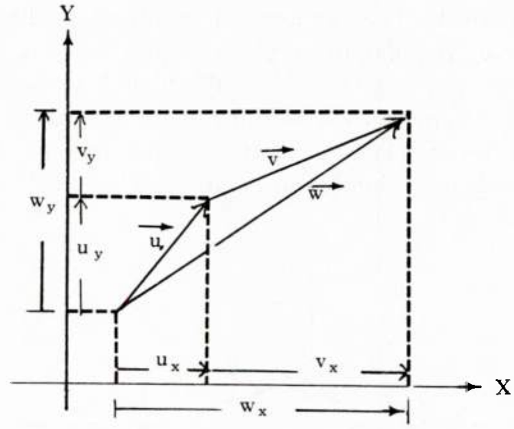
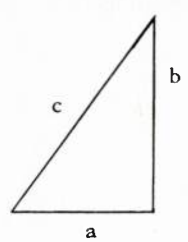


Figure 5. Vector addition in rectangular coordinates.

w_x and w_y are two sides of a right triangle of which the vector w is the hypotenuse. The length of the vector w may be calculated by use of the Pythagorean theorem.

The Pythagorean theorem states that the sum of the squares of the two sides of a right triangle is equal to the square of the hypotenuse (Fig. 6).



$$a^2 + b^2 = c^2$$

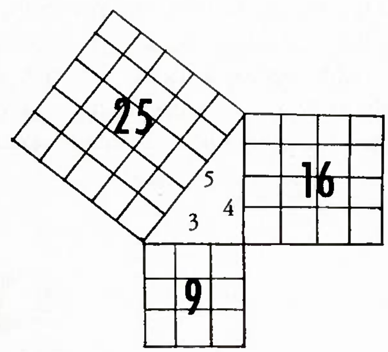


Figure 6. Pythogorean Theorem.

With the help of the Pythagorean theorem, it is easy to derive the equation of a circle whose center is at the origin of a rectangular coordinate system (Fig. 7). From a general point, P(x, y) on the circumference of the circle, draw a line representing the radius to the origin.

Draw another line perpendicular to the x-axis. These two lines form a right triangle whose sides are equal to x and y and whose hypotenuse is equal to R, the radius. If the Pythagorean theorem is applied, the result is $x^2 + y^2 = R^2$. This is the equation of the circle whose center is at the origin of the rectangular coordinate system.

$$x^2 + y^2 = R^2$$

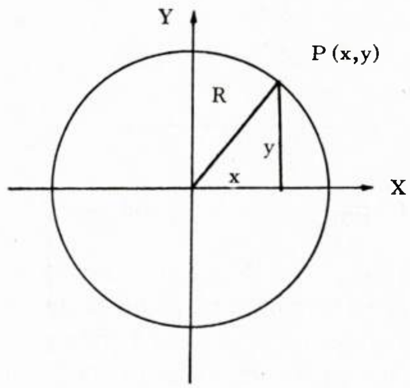


Figure 7. Equation of a circle.

To make the solution of a problem easier, it may be convenient to convert rectangular coordinates to polar coordinates, or vice versa.

Polar Coordinates

Polar coordinates provide another means for locating a point or object on a plane surface. After a pilot tunes in a TACAN station and sees that he is 55 nautical miles (NM) out on the 340° radial, he has identified his position with the use of one form of polar coordinates (Fig. 8). He has specified a distance, 55NM, and an angle of 340°. The TACAN station is assumed to be the pole, or starting point.

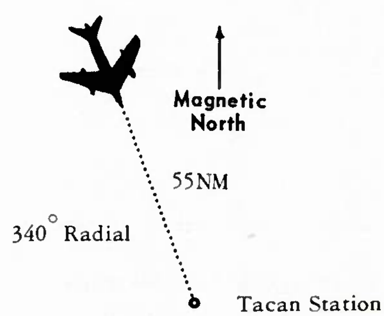


Figure 8. A pilot's position identified by polar coordinates.

In the general mathematical use of polar coordinates, a pole is specified as a starting point. A reference direction called the polar axis is established. The coordinates are given as a radius vector, r, which is a distance, and a vectorial angle, θ, which is measured counterclockwise from the polar axis to the radius vector r, as shown in Figure 9.

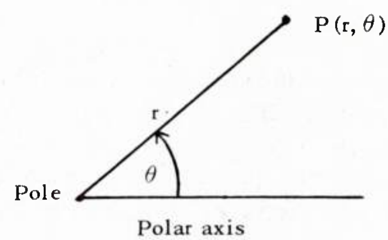


Figure 9. Polar coordinates.

When converting from polar to rectangular coordinates, first superimpose the polar coordinate system on the rectangular coordinate system as shown in Figure 10. The Pythagorean theorem relates the x and y of a given point (P) to r, but more relationships between the rectangular coordinates (x and y) and the polar coordinates (r and θ) are needed. To get these relationships, it is necessary to review some of the basic ideas of trigonometry.

The trigonometric functions are relations between the angles and sides of a right triangle. The sine of the angle θ ($\sin \theta$) is defined as the ratio of the side opposite θ to the hypotenuse, or $\sin \theta = \frac{\text{opp}}{\text{hyp}}$ (Fig. 10).

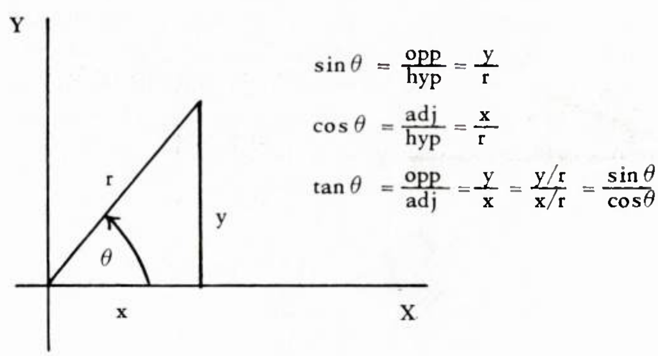
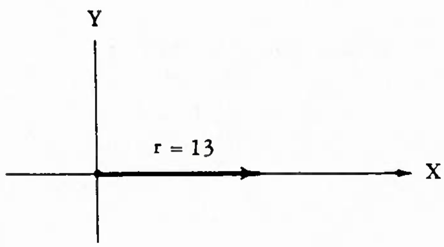


Figure 10. Polar coordinate system superimposed on the rectangular coordinate system.

The cosine of angle θ is the ratio of the side adjacent to θ to the hypotenuse, or $\cos \theta = \frac{\text{adj}}{\text{hyp}}$. The tangent of θ is the ratio of the side opposite to the side adjacent to θ, or $\tan \theta = \frac{\text{opp}}{\text{adj}} = \frac{\sin \theta}{\cos \theta}$.

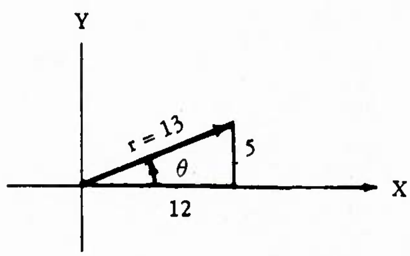
Figure 11 shows how the three functions change in value as the radius r is rotated counterclockwise around the pole and the value of θ varies from 0° to 360°. Values in all four quadrants are shown. A specific value of 13 was assigned to r, and specific values for x and y were chosen so that the sides would be integers. The value of the angle θ was determined from the slide rule. The values of the trigonometric functions can also be found by consulting pages 278 through 283 in *The Engineer's Manual* by Hudson.



$$\sin\theta = \frac{\text{opp}}{\text{hyp}} = \frac{0}{13} = 0 \qquad \theta = 0°$$

$$\cos\theta = \frac{\text{adj}}{\text{hyp}} = \frac{13}{13} = 1$$

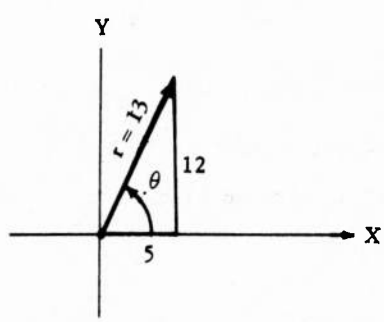
$$\tan\theta = \frac{\text{opp}}{\text{adj}} = \frac{0}{13} = 0$$



$$\sin\theta = \frac{5}{13} = .384 \qquad \theta = 22.6°$$

$$\cos\theta = \frac{12}{13}$$

$$\tan\theta = \frac{5}{12}$$

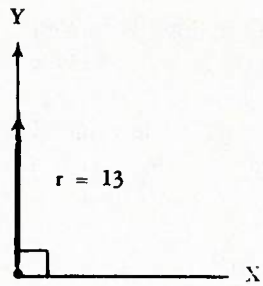


$$\sin\theta = \frac{12}{13} = .923 \qquad \theta = 67.4°$$

$$\cos\theta = \frac{5}{13}$$

$$\tan\theta = \frac{12}{5}$$

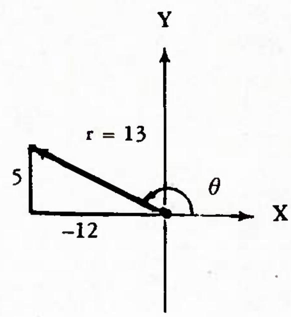
Figure 11. Typical trigonometric values.



$$\sin\theta = \frac{\text{opp}}{\text{hyp}} = \frac{13}{13} = 1 \qquad \theta = 90^\circ$$

$$\cos\theta = \frac{\text{adj}}{\text{hyp}} = \frac{0}{13} = 0$$

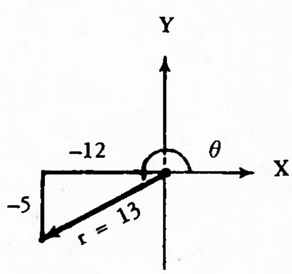
$$\tan\theta = \frac{\text{opp}}{\text{adj}} = \frac{13}{0} = \infty$$



$$\sin\theta = \frac{5}{13} = .384 \qquad \theta = 180\text{–}22.6 = 157.4^\circ$$

$$\cos\theta = -\frac{12}{13}$$

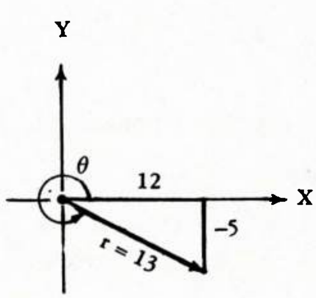
$$\tan\theta = -\frac{5}{12}$$



$$\sin\theta = \frac{-5}{13} = -.384 \qquad \theta = 180 + 22.6 = 202.6^\circ$$

$$\cos\theta = \frac{-12}{13}$$

$$\tan\vartheta = \frac{-5}{-12} = \frac{5}{12}$$



$$\sin\theta = -\frac{5}{13} = -.384 \qquad \theta = 360 - 22.6 = 337.4^\circ$$

$$\cos\theta = \frac{12}{13}$$

$$\tan\theta = -\frac{5}{12}$$

Figure 11. Typical trigonometric values (cont'd).

Figures 12a, b, and c are continuous plots of the three functions, with the numerical values plotted on the vertical, or y-axis and the degrees on the horizontal, or x-axis.

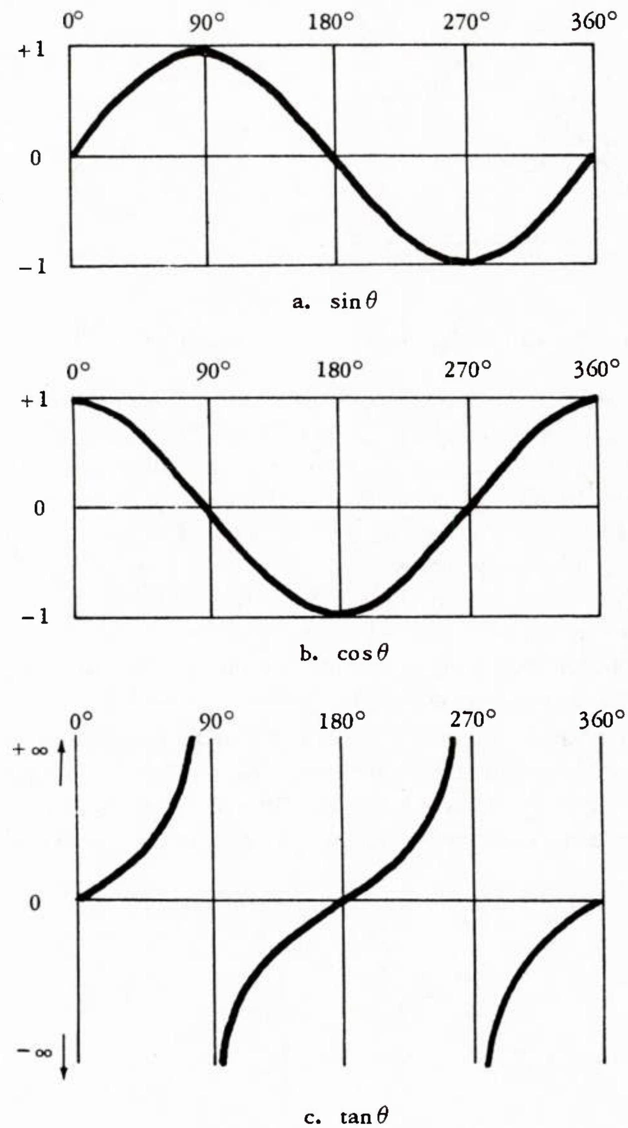
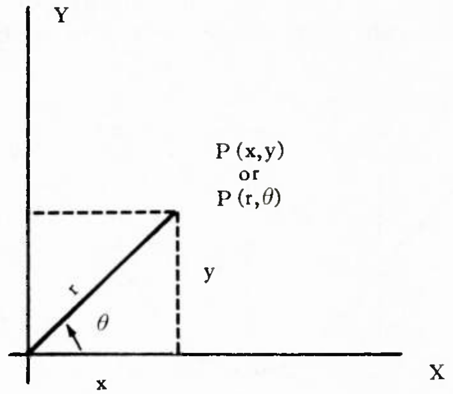


Figure 12. Plots of trigonometric functions.

To convert from polar to rectangular coordinates, simply apply the definitions of the basic trigonometric relationships. Figure 13 shows that $\sin\theta = \frac{\text{opp}}{\text{hyp}} = \frac{y}{r}$. Therefore, $y = r\sin\theta$. In a similar manner, it can be shown that $x = r\cos\theta$. Apply the Pythagorean theorem to arrive at the equation $x^2 + y^2 = r^2$. Finally, by definition, $\tan\theta = \frac{y}{x}$.



$$x^2 + y^2 = r^2$$
$$y = r \sin\theta$$
$$x = r \cos\theta$$
$$\tan\theta = \frac{y}{x}$$

Figure 13. Conversion from polar to rectangular coordinates.

Rectangular coordinates are not necessarily easier to work with. Sometimes it is simpler to use polar coordinates. In general, a straight line adapts itself more readily to rectangular coordinates, and a circle, to polar coordinates. A review of the equations shows why this is true. In rectangular coordinates, $y = mx + b$ is the equation of a straight line. If the conversions $y = r \sin \theta$ and $x = r \cos \theta$ are substituted into the original equation, it becomes $r \sin \theta = mr \cos \theta + b$. This latter equation is still a straight line, but it is more difficult to handle. On the other hand, the equation of a circle in rectangular coordinates is $x^2 + y^2 = R^2$. Therefore, $r^2 = R^2$ or $r = R$ in polar coordinates. Thus, in polar coordinates, the equation of a circle is simpler than it is in rectangular coordinates.

There are three more trigonometric functions that should be mentioned to make the review of trigonometric functions complete. These are the cosecant (csc), the secant (sec), and the cotangent (cot), which are merely the reciprocal of the sine, the cosine, and the tangent, respectively. The last three functions are defined as follows:

$$\csc \theta = \frac{\text{hyp}}{\text{opp}} = \frac{1}{\sin \theta}$$

$$\sec \theta = \frac{\text{hyp}}{\text{adj}} = \frac{1}{\cos \theta}$$

$$\cot \theta = \frac{\text{adj}}{\text{opp}} = \frac{1}{\tan \theta} = \frac{\cos \theta}{\sin \theta}$$

The table below gives the algebraic signs of all six trigonometric functions in the four quadrants.

Functions	Quadrant			
	I	II	III	IV
sin and csc	+	+	−	−
cos and sec	+	−	−	+
tan and cot	+	−	+	−

In numerical computations, it is convenient to relate the trigonometric functions of angles in the second, third, and fourth quadrants to trigonometric functions of angles in the first quadrant since many trigonometric tables only give the functions of angles between 0° and 90°. The most useful such relations are:

If θ is in the second quadrant (90° to 180°)

$$\sin\theta = \sin(180° - \theta)$$
$$\cos\theta = -\cos(180° - \theta)$$

If θ is in the third quadrant (180° to 270°)

$$\sin\theta = -\sin(\theta - 180°)$$
$$\cos\theta = -\cos(\theta - 180°)$$

If θ is in the fourth quadrant (270° to 360°)

$$\sin\theta = -\sin(360° - \theta)$$
$$\cos\theta = \cos(360° - \theta)$$

A highly useful relationship exists between the squares of the sine and of the cosine of any angle.

$$\sin^2\theta + \cos^2\theta = 1$$

This trigonometric identity can be verified by substituting $x = r\cos\theta$ and $y = r\sin\theta$ into the Pythagorean theorem: $x^2 + y^2 = r^2$.

Oblique Triangles

The trigonometric functions of angles are also useful in relating the parts of an oblique triangle. (Note: It is customary to designate the angles of a triangle by the uppercase letters A, B, C, and the side opposite a particular angle by the same lowercase letter as in Figure 14.) In any triangle, the ratio between the sine of an angle and the length of the opposite side is a constant, and is known as the Law of Sines.

Law of Sines $$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$$

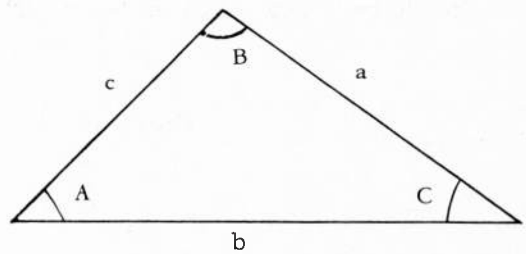


Figure 14. An oblique triangle showing the conventional side and angle notation.

If two angles and a side of any triangle are known, then the remaining parts of the triangle can be determined by the Law of Sines and the relation that the sum of the angles of a triangle is 180°.

$$A + B + C = 180°$$

If two sides and the included angle of any triangle are known, then the third side can be determined by the Law of Cosines:

Law of Cosines

$$a^2 = b^2 + c^2 - 2bc \cos A$$
$$b^2 = a^2 + c^2 - 2ac \cos B$$
$$c^2 = a^2 + b^2 - 2ab \cos C$$

In the special case of a right triangle, it can be seen that the Law of Cosines reduces to the Pythagorean theorem. For example, say C is the right angle. Then $\cos C = \cos 90° = 0$ and $c^2 = a^2 + b^2$.

In summary, if three sides of a triangle are known, or if two sides and the included angle are known, or if a side and any two angles are known, then the other sides and angles can be determined by the Law of Sines and the Law of Cosines.

The application of the principles of trigonometry and algebra allows derivation of the transformations for the conversion of rectangular coordinates to polar coordinates, and vice versa. By means of coordinate transformations the track of a satellite, which is detected and measured with a radar in terms of polar coordinates, can be converted to rectangular grid coordinates for plotting.

Appendix B

DETERMINATION OF THE ANGLE BETWEEN TWO ORBITAL PLANES

FOR SPACE rendezvous to occur both vehicles must simultaneously be in the same orbital plane and be at the same location in identical orbits.[1] For this discussion the orbit requirement will be assumed to have been satisfied independently of the orbital plane requirement. The maneuvering operations necessary to satisfy both requirements may occur in any sequence; however, only the orbital plane requirement will be discussed here.

In space rendezvous operations one encounters the problem of reaching a specified orbital plane from either another orbital plane or a specific launch site. Determination of the plane change angle (α) required to change from one orbital plane to another, at the intersection of the two planes, is complicated only by the requirement to use spherical trigonometry instead of the plane trigonometry to which we are more accustomed.

To simplify the discussion of the solution, the orbital plane of the vehicle is defined as the initial plane, and the desired new orbital plane is defined as the final plane. Since any launch must initially enter an orbital plane which is dependent on the launch site latitude (L), launch azimuth (β), and time and date of launch, the above definitions will suffice.

For the case of an initial plane and a final plane, the solution is dependent on knowing or estimating the right ascension (Ω) and inclination (i) angles of the two planes. For the case of a final plane and a launch site, the solution is dependent on: the Ω and i angles for the plane; time between launch and passage of the orbital plane over the launch site; and the launch latitude and azimuth.

Two examples below illustrate the mathematics involved and the approximation techniques available for obtaining solutions. In the first example it will be shown that only the difference in right ascension angles ($\Delta\Omega$) is required; the specific values of Ω need not be known.

Case I

A vehicle in orbit 1 is to be maneuvered into the plane of orbit 2. To accomplish this maneuver the value of the plane change angle is required.

Orbital Parameters

	Orbit 1	Orbit 2
Ω	30°	45°
i	28°	30°

[1] Identical orbits occur when the magnitude and orientation of the major axes (2a) are the same and the eccentricities (ϵ) are equal.

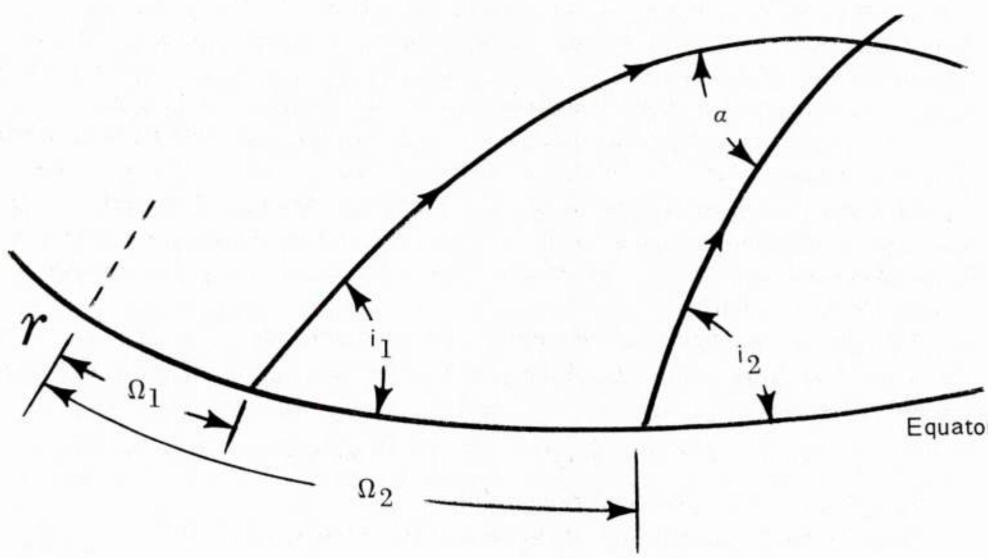
From spherical trigonometry the plane change angle can be determined from:

$$\cos\left(\frac{\alpha}{2}\right) = \frac{\cos\frac{i_2 - i_1}{2}}{\cos(x)} \cos\frac{\Omega_2 - \Omega_1}{2}$$

Where the angle x, a dummy angle used in the calculation, is determined from:

$$\tan(x) = \frac{\cos\frac{i_2 + i_1}{2}}{\cos\frac{i_2 - i_1}{2}} \tan\frac{\Omega_2 - \Omega_1}{2}$$

The geometry of the problem is as shown below.



From the equations and the geometry it is apparent that: only the difference ($\Delta\Omega$) in Ω is required; it will always be positive; and it can be reduced to an angle between 0 and 180 degrees. Therefore, for this problem $\Delta\Omega$ is 15°.

Substituting to determine the angle x:

$$\tan x = \frac{\cos\frac{30^\circ + 28^\circ}{2}}{\cos\frac{30 - 28}{2}} \tan\frac{15^\circ}{2} = \frac{\cos(29^\circ)}{\cos(1^\circ)} \tan(7.5^\circ)$$

$$= \frac{0.8746197}{0.9998477} \times (0.131652) = 0.1151629$$

Therefore, $x = 6^\circ\ 34'$ and $\cos x = 0.9934727$

Substituting to determine the plane change angle α:

$$\cos\left(\frac{\alpha}{2}\right) = \frac{\cos \dfrac{30^\circ - 28^\circ}{2}}{0.9934727} \cos(7.5^\circ)$$

$$= \frac{0.9998477}{0.9934727} \times (0.9914449) = 0.997807$$

Therefore, $\frac{\alpha}{2} = 3^\circ\ 48'$ or $\alpha = 7^\circ\ 36'$.

The required plane change angle is $7^\circ\ 36'$.

Case II

A spacecraft on a launch vehicle at Cape Kennedy ($L = 28^\circ$) must attain orbit in a plane which has a southwest-northeast orientation and an inclination angle of 30°. Ideally, the vehicle would be launched and injected along the required azimuth angle, when the desired orbital plane was directly over the launch site. This is, however, a very difficult timing problem. The spacecraft/launch vehicle combination has limited velocity change capabilities which narrows the "launch window." A launch occurring within this window of time will permit injection into an initial parking orbit and a subsequent plane change maneuver to the desired orbital plane.

Because of launch delays the launch site passed through the desired final orbital plane 8 minutes prior to lift-off. Since the launch vehicle guidance was programmed prior to launch, the spacecraft will be injected into an initial parking orbit, and then it must make a plane change to the desired final orbital plane. The programmed launch azimuth angle is 79°. What is the value of the plane change angle required to complete the mission?

The apparent unknowns are the initial orbit inclination angle, and the $\Delta\Omega$.

To find the inclination angle of this parking orbit, use the equation for inclination angle of the launch:

$$\cos i = \cos(L) \sin(\beta) = \cos(28^\circ) \sin(79^\circ) = 0.8660$$

Therefore, $i = 30^\circ$ for the parking orbit.

It is possible to approximate $\Delta\Omega$ using the earth's rate of rotation.

$$\text{Rotation Rate} = \frac{360^\circ}{24 \text{ hrs}} = \frac{360^\circ}{(24 \times 60) \text{ min}} = 0.25^\circ/\text{min}$$

For the situation of equal inclination angles, the $\Delta\Omega$ is equal to the time between launch and passage of the final orbital plane overhead, multiplied by the earth's rotation rate:[2]

$$\Delta\Omega = 0.25^\circ/\text{min} \times 8 \text{ min} = 2^\circ$$

By making use of the same equations used for Case I, the plane change angle α will be approximately 1°.

---

[2] For the case of non equal inclination angles it is necessary to add a term $\Delta\Omega^*$.

$$\Delta\Omega^* = \Omega^*_1 \pm \Omega^*_2 \quad \begin{cases} - & \text{Both inclination angles in the same quadrant} \\ + & \text{Inclination angles in different quadrants} \end{cases}$$

$\Delta\Omega^*$ must be positive
To find the Ω^*'s use:

$\sin \Omega^*_1 = \tan(L) \cot(i_1)$

$\sin \Omega^*_2 = \tan(L) \cot(i_2)$

Then the total $\Delta\Omega$ equals the $\Delta\Omega$ due to the earth's rotation plus the $\Delta\Omega^*$ due to the difference in inclination angles.

List of Symbols

α—Plane Change Angle
β—Azimuth Angle
$\Delta\Omega$—Difference in Right Ascension Angles
Ω—Right Ascension Angle
i—Inclination Angle
L—Angle of Latitude
x—Dummy Angle

Subscripts

1. Left hand orbital plane
2. Right hand orbital plane

Superscripts

* Dummy Variable

Appendix C

BALLISTIC MISSILE TRAJECTORIES

THE PURPOSE of this section is to present the quantitative analysis and geometry of ballistic missile trajectories. The analysis is based upon the two-body problem with the earth as the large, central body. Specific problems concern the range of a ballistic missile for specified burnout conditions and the effect of errors in burnout conditions on the desired impact point.

In order to simplify the problem so that the fundamentals are highlighted, a series of practical assumptions are made. The earth is assumed to be spherical and non-rotating. The effect of the earth's oblateness is very complicated; an extensive study is beyond the scope of this text. There is an oblateness effect upon the trajectory itself; the target, therefore, is not in the same position that it would be on a spherical earth. The effect of the earth's oblateness is more pronounced upon satellite trajectories than upon ballistic missile trajectories because of the longer time of flight. The powered and reentry portions of an actual trajectory are not considered here; the two-body problem cannot be applied to these parts of the trajectory, as energy is dissipated in air friction in both parts, and propulsive energy is being added to the system during the powered trajectory; thus, the conditions of constant mechanical energy and constant angular momentum do not apply during these portions of the trajectory.

Geometry and Standard Values of Earth Parameters

Before proceeding to computational analysis, it is essential that the geometry of the ballistic missile problem be understood. The geometry is presented in Fig. 1. The symbols are defined below. Numerical calculations require the use of earth parameters, and standard values of these are also presented below.

- r_{bo} radius to the burnout point
- v_{bo} magnitude of the velocity (speed) at the burnout point
- ϕ_{bo} angle between the velocity vector (tangent to the trajectory) and the local horizontal at the burnout point
- θ angle from the point of apogee, measured in the direction of motion
- Λ total range angle
- Γ powered trajectory range angle
- Ω reentry range angle
- Ψ free-flight range angle
- r_e radius of the earth $= 3440$ NM $= 20.9 \times 10^6$ feet
- g_o acceleration of gravity at sea level $= 32.2 \dfrac{\text{ft}}{\text{sec}^2}$
- μ defined as Gm_1, where G is the Universal Gravitational Constant and m_1 is the mass of the earth, $\mu = 14.08 \times 10^{15} \dfrac{\text{ft}^3}{\text{sec}^2}$

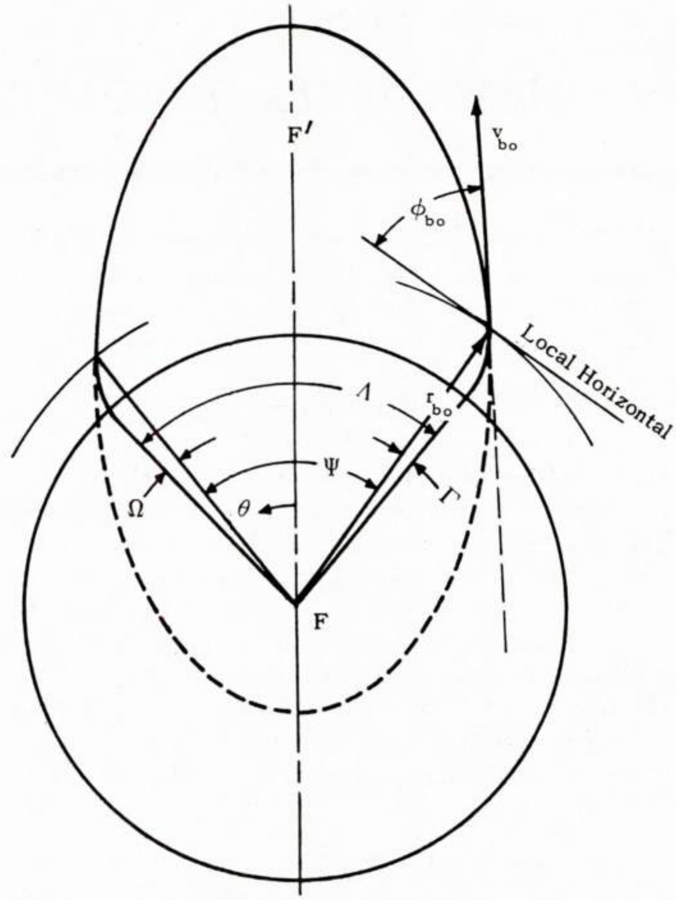


Figure 1. The ballistic missile trajectory.

General Ballistic Missile Problem

If a launch point and target for a ballistic missile are selected, then only the range is known. If the burnout conditions of v_{bo}, r_{bo} and ϕ_{bo} are given, the trajectory equation is uniquely determined, and a definite range will be obtained. Actually, there are many possible combinations of these four parameters, and no good basis for a choice exists until some other condition is imposed. One such condition might be to specify that v_{bo} be the smallest that will get the payload to the target. The converse of this statement is to require maximum range for a given v_{bo}. Clearly, a relationship between range and the trajectory parameters is needed and must be found.

The Range Equation

Symmetrical trajectories ($h_{bo} = h_{re}$) involve simpler mathematics than do unsymmetrical trajectories ($h_{bo} \neq h_{re}$). Because the symmetrical trajectories adequately demonstrate the principles involved, this treatment of ballistic missiles will be limited to symmetrical trajectories. By using the polar equation for the two-body

trajectory and suitable geometric substitutions, the ballistic missile range equation can be developed in terms of Ψ (the free flight range angle) and the burnout conditions (v_{bo}, r_{bo}, and ϕ_{bo}). It is given here as:

$$\cot \frac{\Psi}{2} = \frac{2\mu}{v^2_{bo}\, r_{bo}} \csc 2\phi_{bo} - \cot \phi_{bo} \qquad (1)$$

Equation (1) can be simplified by substitution of the following convenient definition:

$$\text{let } Q = \frac{v^2 r}{\mu} \qquad (2)$$

$$\text{then } Q_{bo} = \frac{v^2_{bo}\, r_{bo}}{\mu}$$

The parameter Q has interesting characteristics. It is equal to 1 for circular orbital conditions, and it is equal to 2 for minimum escape conditions. With this substitution, equation (1) becomes

$$\cot \frac{\Psi}{2} = \frac{2}{Q_{bo}} \csc 2\phi_{bo} - \cot \phi_{bo} \qquad (3)$$

The following example problem will illustrate how (1) may be used.

Example Problem

The following information is given about a ballistic missile:

$$v_{bo} = 16{,}000 \text{ ft/sec}$$
$$\phi_{bo} = 21°$$
$$h_{bo} = 164.5 \text{ NM}$$

What is the free-flight ground range of this missile on a spherical, non-rotating earth?

Problem Solution

Seven graphs, Figs 4-10, are included at the end of this section for use in graphical solutions.

Graphical: Graph Fig. 4, $Q_{bo} = .4$

Graph Fig. 7, $R_{ff} = 1445$ NM

Calculated:

$$Q_{bo} = \frac{v_{bo}^2\, r_{bo}}{\mu} = \frac{(1.6 \times 10^4 \text{ ft/sec})^2 \times 21.9 \times 10^6 \text{ ft}}{14.08 \times 10^{15} \text{ ft}^3/\text{sec}^2}$$

$$Q_{bo} = .398 \approx .4$$

$$\cot \frac{\Psi}{2} = \frac{2}{Q_{bo}} \csc 2\phi_{bo} - \cot \phi_{bo} = \frac{2}{.4} \csc 42° - \cot 21°$$

$$\cot \frac{\Psi}{2} = \frac{2 \times 1.49}{.4} - 2.6 = 4.84$$

$$\frac{\Psi}{2} = 11.7°$$

$$\Psi = 23.4°$$

$$R_{ff} = 23.4° \times 60 \text{ NM/°} = 1404 \text{ NM}$$

The Equation for Flight Path Angle

A more practical ballistic missile problem results from defining a launch point, target, and missile. The launch point and target determine the range (Ψ) while the missile is capable of attaining a certain burnout velocity and height. In this problem, then, Ψ, v_{bo}, and r_{bo} are given and it is desired to find ϕ_{bo}. Equation (1) may be solved for ϕ and be refined to the following form:

$$\cot \phi_{bo} = \frac{-\cot \frac{\Psi}{2} \pm \sqrt{\cot^2 \frac{\Psi}{2} + \frac{4}{Q_{bo}}\left(1 - \frac{1}{Q_{bo}}\right)}}{2\left(1 - \frac{1}{Q_{bo}}\right)} \qquad (4)$$

Equation (4) points out some very important facts. When $Q_{bo} < 1$, Ψ must be less than 180°, and there are two values of ϕ_{bo} for a given range, v_{bo}, and r_{bo}. There are two trajectories, then, to the target. The trajectory corresponding to the larger value of flight path angle is called the "high" trajectory; the trajectory associated with the smaller flight path angle is the "low" trajectory.

A simple illustration of the principle involved is the behavior of water discharged from a garden hose. With a constant water pressure and nozzle setting, the velocity of the water leaving the nozzle is fixed. If a target well within the reach of the water is selected, the target can be hit by both a high and a low trajectory.

The high and low trajectories are sometimes referred to as "lofted" and "flat" trajectories, respectively.

When $1 \leq Q_{bo} < 2$, and $\Psi < 180°$, there is still a high trajectory, but the low trajectory does not exist because it would penetrate the earth. Interestingly enough, however, when $1 < Q_{bo} < 2$, Ψ can be greater than 180°. An illustration of such a trajectory would be a missile directed at the Asian continent from North America via the south polar region. Although such a trajectory would be capable of avoiding a northern radar "fence," it would be costly in terms of payload delivered.

It is interesting to compare parameters of the high and low trajectories. As v_{bo} and r_{bo} are the same for the two trajectories, E is also the same. But $a = -\mu/2E$, so the major axes of the trajectories are the same. The angular momentum is smaller for the high trajectory; c and ϵ are larger. It is also evident that the time-of-flight on the high trajectory will be longer.

For the constraint of fixed energy, it is reasonable to anticipate that there are certain ranges which are attainable and certain ranges which are not attainable; between these two groups there is the limiting case in which the range is just barely attainable.

The usual way to find such a maximum is to solve for the dependent variable (in this case the range, Ψ) in terms of the independent variable (in this case, ϕ_{bo}). The partial derivative of the dependent variable with respect to the independent variable is then formed and set equal to zero to determine the conditions for a maximum (or a minimum). In our case the range equation expresses Ψ in terms of ϕ_{bo} and the partial of Ψ with respect to ϕ_{bo} when set equal to zero is:

$$\frac{\partial \Psi}{\partial \phi_{bo}} = \frac{2 \sin (\Psi + 2\phi_{bo})}{\sin 2\phi_{bo}} - 2 = 0$$

Hence

$$\sin(\Psi + 2\phi_{bo}) - \sin 2\phi_{bo} = 0 \tag{5}$$

Using the trigonometric identity

$$\sin x - \sin y = 2 \sin \tfrac{1}{2}(x - y) \cos \tfrac{1}{2}(x + y)$$

equation (5) becomes

$$2 \sin \Psi/2 \cos \tfrac{1}{2}(\Psi + 4\phi_{bo}) = 0 \tag{6}$$

Equation (6) expresses the conditions under which a maximum or a minimum will occur. One condition occurs when $\sin \Psi/2 = 0$. This implies that Ψ is either 0 or 2π, neither of which is of interest for a ballistic missile. The case of interest occurs when

$$\cos \tfrac{1}{2}(\Psi + 4\phi_{bo}) = 0$$

or

$$\tfrac{1}{2}(\Psi + 4\phi_{bo}) = \pi/2$$

from which

$$\phi_{bo} = \tfrac{1}{4}(\pi - \Psi) \tag{7}$$

Equation (7) states the relationship which must exist between Ψ and ϕ_{bo} on a maximum range trajectory. It is important to observe that there is one and only one value of ϕ_{bo} for a maximum range trajectory; the maximum range trajectory is unique.

ECCENTRICITY EQUATION.—It is inconvenient to calculate ϵ from the equation, $\epsilon = \sqrt{1 + \frac{2EH^2}{\mu^2}}$, when given the range and burnout conditions of a ballistic missile. Eccentricity, as a function of range and burnout conditions, would enable more rapid calculations. Two forms of the eccentricity equation are given here as a convenience:

$$\epsilon = \frac{\sin \phi_{bo}}{\sin\left(\phi_{bo} + \frac{\Psi}{2}\right)} \tag{8}$$

$$\epsilon = \sqrt{Q^2_{bo} \cos^2 \phi_{bo} - 2Q_{bo} \cos^2 \phi_{bo} + 1} \tag{9}$$

Range and Azimuth Errors

Still considering the earth to be non-rotating, the target, launch (or burnout) point, and center of the earth define a plane in which the trajectory, and thus the velocity vector, must lie. The burnout velocity vector can have errors in the intended plane as follows: position, magnitude, and flight path angle. It is also possible for the burnout velocity vector to have errors out of the intended plane in which case a new plane is defined. The most important error out of the intended plane is azimuth error. It will be considered before the errors in the intended plane, as the treatment can be brief.

AZIMUTH ERROR.—Figure 2 illustrates the geometry of azimuth error. The arc length, a, represents the cross-range error. The arc lengths, b and c, are equal to each other and are the range. It is usual in spherical trigonometry to measure arc lengths in terms of the angle subtended at the center of the sphere. Using this convention, $b = c = \Psi$, the range angle defined earlier. The law of cosines in spherical trigonometry is:

$$\cos a = \cos b \cos c + \sin b \sin c \cos A$$

which becomes:

$$\cos a = \cos^2 \Psi + \sin^2 \Psi \cos A$$

As A is a small angle, cos A can be approximated by $1 - \frac{A^2}{2}$; then,

$$\cos a = \cos^2 \Psi + \sin^2 \Psi - \frac{A^2 \sin^2 \Psi}{2} = 1 - \frac{A^2 \sin^2 \Psi}{2}$$

But a is also a small angle so that cos a can be approximated by $1 - \frac{a^2}{2}$; substituting this,

$$1 - \frac{a^2}{2} = 1 - \frac{A^2 \sin^2 \Psi}{2}$$

And:

$$a = A \sin \Psi \tag{10}$$

Equation (10) is in terms of angular measure; in order to have the cross-range error in linear dimensions, the radius of the earth must be multiplied by the central angle, a, if it is in radians; if it is in degrees, the relationship that 60NM on the surface are equivalent to a 1 degree central angle at the center of the earth may be used.

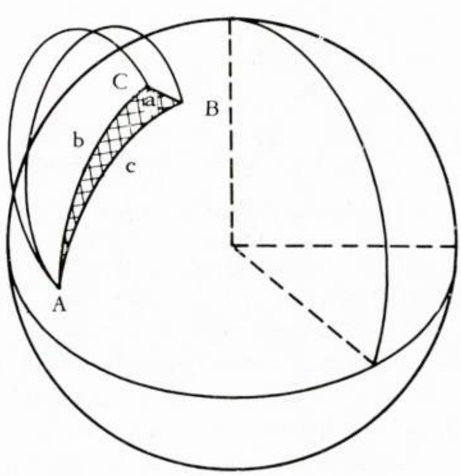


Figure 2. Geometry of azimuth error.

quantity $\frac{\partial \Psi_{bo}}{\partial \phi_{bo}}$. Thus, the effects of each of the possible contributors to free flight range error are influenced by the factors $\frac{\partial \Psi}{\partial v_{bo}}$, $\frac{\partial \Psi}{\partial r_{bo}}$, and $\frac{\partial \Psi}{\partial \phi_{bo}}$. They are referred to as "influence coefficients," since these factors exert an influence on the size of the range error resulting from burnout errors.

There are three types of problems that can be defined in terms of the influence coefficients and equation (12). (1) Given the guidance system capabilities [Δv_{bo}, Δr_{bo} and $\Delta \phi_{bo}$] and the influence coefficients, find the range error. (2) Given the allowable range error [$\Delta \Psi$] and the influence coefficients, find the allowable burnout errors [Δv_{bo}, Δr_{bo}, $\Delta \phi_{bo}$]. (3) Given the allowable range error [$\Delta \Psi$] and the guidance system capabilities [Δv_{bo}, Δr_{bo}, $\Delta \phi_{bo}$], find a trajectory which yields an appropriate set of influence coefficients. Type 3 is by far the most difficult problem since the first two types imply that the trajectory is known. In any case, inseparably related to the study of range errors is the evaluation of the influence coefficients. All of the influence coefficients are obtained by differentiating the range equation:

$$\cot \frac{\Psi}{2} = \frac{2\mu}{v^2_{bo}\, r_{bo}} \csc 2\phi_{bo} - \cot \phi_{bo}$$

In particular the influence coefficient for burnout flight path angle is obtained by taking the partial derivative with respect to ϕ_{bo}.

$$\frac{\partial \Psi}{\partial \phi_{bo}} = \frac{2 \sin (\Psi + 2\phi_{bo})}{\sin 2\phi_{bo}} - 2 \qquad (13)$$

Thus, range and flight path angle determine the influence coefficient for errors in flight path angle. Figure 8 is a plot of this influence coefficient for specific values of Q at various free flight ranges. For convenience in computation the values of the influence coefficient are plotted in *NM* range error per degree error in the flight path angle alignment.

The burnout speed error influence coefficient is obtained by differentiating the range equation with respect to v_{bo}:

$$\frac{\partial \Psi}{\partial v_{bo}} = \frac{8\mu}{v^3_{bo}\, r_{bo}} \frac{\sin^2 \frac{\Psi}{2}}{\sin 2\phi_{bo}} \qquad (14)$$

For any range, and ϕ_{bo} less than 90°, $\frac{\partial \Psi}{\partial v_{bo}}$ will be positive. Hence, a positive speed error will increase the range. If the actual burnout speed is greater than the intended value, the missile will overshoot the target. Conversely, if the actual burnout speed is less than the intended value, the missile will always fall short of the target.

The influence coefficient is determined by the trajectory requirements of range, v_{bo}, r_{bo}, and ϕ_{bo}. Δv is the factor controlled by the guidance system. Figure 9 is a plot of this influence coefficient for specific values of Q at various free flight ranges. For convenience in computation the values of the influence coefficient are plotted in *NM* range error *per ft/sec* error in the velocity magnitude at burnout.

The influence coefficient is smaller for the high trajectory resulting in less error at the target for a given speed error. Some missiles have guidance systems which are accurate enough to permit use of the low trajectory at ICBM range.

From equation (10), it can be seen that, when the range angle, Ψ, is 90 degrees, maximum error occurs. Also, when Ψ is 180 degrees, the azimuth error returns to zero.

RANGE ERROR AS A FUNCTION OF ERRORS IN THE TRAJECTORY PLANE.—In this section it is assumed that the burnout velocity vector lies in the intended trajectory plane (hence the cross-range error is zero), but the possibility exists that the velocity vector may have errors in position, magnitude and flight path angle. To trace the possible sources of range errors, it is helpful to write down the equation for the total range angle (Λ) in terms of the powered range angle, (Γ), the free flight range angle (Ψ), and the reentry range angle (Ω). This equation is simply:

$$\Lambda = \Gamma + \Psi + \Omega$$

Since Λ is the sum of its three parts, any change in the total range angle ($\Delta\Lambda$) can be traced to three possible sources. That is:

$$\Delta\Lambda = \Delta\Gamma + \Delta\Psi + \Delta\Omega \tag{11}$$

Since the changes ($\Delta\Lambda$, $\Delta\Gamma$, $\Delta\Psi$, $\Delta\Omega$) in the last equation represent departures from the desired trajectory, they may properly be termed range errors. An error committed during any portion of the trajectory contributes directly to the total range error $\Delta\Lambda$. The three possible range errors are not independent, but each contributes to the total range error, and none can be ignored.

Concentrate attention upon the free flight range error $\Delta\Psi$ and trace the possible source of $\Delta\Psi$. The range equation, equation (1), expresses the free flight range as a function of the burnout conditions, v_{bo}, r_{bo}, and ϕ_{bo}. Following the rules for the formation of the total differential of an explicit function:

$$d\Psi = \frac{\partial\Psi}{\partial v_{bo}} dv_{bo} + \frac{\partial\Psi}{\partial r_{bo}} dr_{bo} + \frac{\partial\Psi}{\partial\phi_{bo}} \phi_{bo}$$

This is an equation for the total differential, $d\Psi$. The equation for the change in free flight range, $\Delta\Psi$, reads:

$$\Delta\Psi \cong \frac{\partial\Psi}{\partial v_{bo}} \Delta v_{bo} + \frac{\partial\Psi}{\partial r_{bo}} \Delta r_{bo} + \frac{\partial\Psi}{\partial\phi_{bo}} \Delta\phi_{bo} \tag{12}$$

$\Delta\Psi$ is the free flight range error. The quantities, Δ_{bo}, Δr_{bo}, $\Delta\phi_{bo}$ represent departure from the desired burnout conditions and so represent the burnout speed error, the burnout altitude error, and the burnout flight path angle error, respectively.

Next, consider the three quantities $\frac{\partial\Psi}{\partial v_{bo}}, \frac{\partial\Psi}{\partial r_{bo}}$, and $\frac{\partial\Psi}{\partial\phi_{bo}}$. In equation (11) each of the possible errors enters the summation in a one-to-one manner while in equation (12) the possible errors are multiplied by the factors $\frac{\partial\Psi}{\partial v_{bo}}, \frac{\partial\Psi}{\partial r_{bo}}$, and $\frac{\partial\Psi}{\partial\phi_{bo}}$. This fundamental difference means, for example, that a burnout flight path angle error of one degree will not necessarily produce a range error of one degree but will produce an error which depends upon the size of the

The last influence coefficient of interest related errors in altitude (radius) to the resulting range errors. Here the missile is considered to have attained the proper burnout velocity on the intended radial, but at an incorrect radius.

$$\frac{\partial \Psi}{\partial r_{bo}} = \frac{4\mu}{v^2_{bo}\, r^2_{bo}} \frac{\sin^2 \dfrac{\Psi}{2}}{\sin 2\phi_{bo}} \tag{15}$$

Figure 10 is a plot of this influence coefficient for specific values of Q at various flight ranges. For convenience in computation the values of the influence coefficient are plotted in *NM* range error *per 10^3 ft* error in radius (or altitude) at burnout.

Again, the error is not as great for the high trajectory. The range error resulting from an error in burnout height should not be confused with the change in range which would be obtained in launching a missile from a high elevation. If an Atlas, say, were launched from the top of Pikes Peak, it would have a range increase due to $\frac{\partial \Psi}{\partial r_{bo}}$, but it would have another, probably greater, range increase resulting from its powered trajectory occurring in less dense atmosphere. The same missile launched at sea level would have to expend considerable energy overcoming the large drag forces up to 14,000 ft. The effect of a launch altitude then is primarily related to the powered trajectory and is a separate phenomenon from the influence coefficient effect.

Example Problem

Telementry from Cape Kennedy indicated that an ICBM achieved burnout conditions of h_{bo} = 164.5 NM, v_{bo} = 22,700 ft/sec, and $\phi_{bo} = 30°$.

(a) The particular guidance system used had an accuracy of $\Delta h_{bo} = \pm 1000$ ft, $\Delta v_{bo} = \pm 2$ ft/sec, and $\Delta \phi_{bo} = \pm .2°$. ($3.49 \times 10^{-3}$ radians). What is the total free-flight range error? *Ans.* Q = .8, R_{ff} = 4900 NM; $\Delta\Psi = -3.99$ NM error.

(b) An improved guidance system which is able to control $\Delta\phi = \pm .05°$ (8.72×10^{-4} radians) is installed. What is the total free-flight error now? *Ans.* $\Delta\Psi = .51$ NM error.

(c) Assuming Δh_{bo} and $\Delta \phi_{bo}$ are both positive, what correction in v_{bo} (Δv_{bo}) would be necessary to zero the free-flight range error found in (b) above? *Ans.* Δv_{bo} = 1.4 ft/sec.

Problem Solution

a. *Graphical:* Graph Fig. 5, Q = .8 (by interpolation) *Ans.*

Graph Fig. 7, R_{ff} = 4900 NM (High trajectory) *Ans.*

Graph Fig. 8, $\frac{\partial \Psi}{\partial \phi_{bo}} = -30$ NM/deg

Graph Fig. 9, $\frac{\partial \Psi}{\partial v_{bo}} = 0.8$ NM/ft/sec

Graph Fig. 10, $\frac{\partial \Psi}{\partial r_{bo}} = .41$ NM/10^3 ft

$$\Delta\Psi = \left(\frac{\partial \Psi}{\partial \phi_{bo}} \Delta\phi_{bo}\right) + \left(\frac{\partial \Psi}{\partial v_{bo}} \Delta v_{bo}\right) + \left(\frac{\partial \Psi}{\partial r_{bo}} \Delta r_{bo}\right)$$

Arbitrarily assuming all injection errors are positive:

$$\Delta\Psi = \left(-30 \frac{\text{NM}}{\text{deg}} \times 0.2 \text{ deg}\right) + \left(0.8 \frac{\text{NM}}{\text{ft/sec}} \times 2 \text{ ft/sec}\right) + \left(.41 \frac{\text{NM}}{10^3 \text{ ft}} \times 10^3 \text{ ft}\right)$$

$$\Delta\Psi = (-6.0 \text{ NM} + 1.6 \text{ NM} + .41 \text{ NM}) = -4.0 \text{ NM} \qquad \textit{Ans.}$$

a. *Calculated:*

$$Q = \frac{v_{bo}^2 r_{bo}}{\mu} = \frac{(2.27 \times 10^4 \text{ ft/sec})^2 (20.9 \times 10^6 + 164.5 \times 6080) \text{ ft}}{14.08 \times 10^{15} \text{ ft}^3/\text{sec}^2}$$

$$Q = \frac{(5.14 \times 10^8)(21.9 \times 10^6)}{14.08 \times 10^{15}} = .8 \qquad \textit{Ans.}$$

$$\cot \frac{\Psi}{2} = \frac{2}{Q_{bo}} \csc 2\phi_{bo} - \cot \phi_{bo} = \frac{2}{.8} \csc 60° - \cot 30°$$

$$\cot \frac{\Psi}{2} = 2.5\,(1.155) - 1.732 = 2.888 - 1.732 = 1.156$$

$$\frac{\Psi}{2} = 40.8° \qquad \Psi = 81.6°$$

$$R_{ff} = 81.6° \times 60 \text{ NM/}° = 4896 \text{ NM} \qquad \textit{Ans.}$$

$$\frac{\partial\Psi}{\partial\phi_{bo}} = \frac{2 \sin (\Psi + 2\phi_{bo})}{\sin 2\phi_{bo}} - 2 = \frac{2 \sin (81.6° + 60°)}{\sin 60°} - 2$$

$$\frac{\partial\Psi}{\partial\phi_{bo}} = \frac{2 \sin 141.6°}{\sin 60°} - 2 = \frac{2 \sin 38.4°}{\sin 60°} - 2 = \frac{2\,(.621)}{.866} - 2$$

$$\frac{\partial\Psi}{\partial\phi_{bo}} = \frac{1.242}{.866} - 2 = 1.434 - 2 = -.566$$

$$\frac{\partial\Psi}{\partial v_{bo}} = \frac{8\mu}{v_{bo}^3 r_{bo}} \frac{\sin^2 \frac{\Psi}{2}}{\sin 2\phi_{bo}} \qquad \frac{8 \times 14.08 \times 10^{15}}{11.70 \times 10^{12} \times 21.9 \times 10^6} \times \frac{.653^2}{.866}$$

$$\frac{\partial\Psi}{\partial v_{bo}} = \left(\frac{112.6 \times 10^{-3}}{256.2}\right)(.492) = (.44 \times 10^{-3})(.492) = 2.16 \times 10^{-4} \frac{\text{rad}}{\text{ft/sec}}$$

$$\frac{\partial\Psi}{\partial r_{bo}} = \frac{4\mu}{v_{bo}^2 r_{bo}^2} \frac{\sin^2 \frac{\Psi}{2}}{\sin 2\phi_{bo}} = \frac{4 \times 14.08 \times 10^{15}}{5.15 \times 10^8 \times 4.80 \times 10^{14}} \times \frac{.653^2}{.866}$$

$$\frac{\partial\Psi}{\partial r_{bo}} = \frac{56.3 \times 10^{-7}}{24.72} (.492) = (2.28 \times 10^{-7})(.492) = 11.2 \times 10^{-8} \text{ rad/ft}$$

$$\Delta\Psi = \left(\frac{\partial\Psi}{\partial\phi_{bo}}\Delta\phi_{bo}\right) + \left(\frac{\partial\Psi}{\partial v_{bo}}\Delta v_{bo}\right) + \left(\frac{\partial\Psi}{\partial r_{bo}}\Delta r_{bo}\right)$$

$$\Delta\Psi = (-.566)\left(\frac{.2^\circ}{57.3^\circ/\text{rad}}\right) + \left(2.16 \times 10^{-4}\ \frac{\text{rad}}{\text{ft/sec}}\right)(2\ \text{ft/sec}) + (11.2 \times 10^{-8}\ \text{rad/ft} \times 1000\ \text{ft})$$

$$\Delta\Psi = -.00198 + .000432 + .000112 = -.1436\ \text{rad}$$

$$\Delta\Psi = -.001436\ \text{rad} + 3440\ \frac{\text{NM}}{\text{rad}} = -4.94\ \text{NM}$$ *Ans.*

b. $$\Delta\Psi = \left(-30\ \frac{\text{NM}}{\text{deg}} \times .05\ \text{deg}\right) + 1.6\ \text{NM} + .41\ \text{NM} = -1.5\ \text{NM}$$

$$+ 1.6\ \text{NM} + .41\ \text{NM}$$

$$\Delta\Psi = .51\ \text{NM}$$ *Ans.*

c. $$\Delta\Psi = 0 = \left(\frac{\partial\Psi}{\partial\phi_{bo}}\Delta\phi_{bo}\right) + \left(\frac{\partial\Psi}{\partial v_{bo}}\Delta v_{bo}\right) + \left(\frac{\partial\Psi}{\partial r_{bo}}\Delta r_{bo}\right)$$

$$-\frac{\partial\Psi}{\partial v_{bo}}\Delta v_{bo} = -1.5\ \text{NM} + .41\ \text{NM} = -1.1\ \text{NM}$$

$$\Delta v_{bo} = \frac{-1.1\ \text{NM}}{-.8\ \dfrac{\text{NM}}{\text{ft/sec}}} = 1.4\ \frac{\text{ft}}{\text{sec}}$$ *Ans.*

Time of Flight of Ballistic Missiles

In choosing the trajectory for a ballistic missile the time of flight of the missile is an important consideration. The range from launch point to target on a rotating earth is affected by the time of flight. The time available for interception or other reaction depends upon the time of flight. The time of flight during free flight is of primary concern at this point.

The general time of flight methods are applicable to any ballistic missile trajectory, but there are certain cases in which the formulas and computations are particularly simple.

In Appendix D there is derived a formula for the time of flight of a ballistic missile for which the burnout and reentry heights are equal. It is given as:

$$t_\psi = 2\sqrt{\frac{a^3}{\mu}}\,(\pi - u_{bo} + \epsilon \sin u_{bo}) \qquad (16)$$

In Fig. 3 the time of flight on symmetrical trajectories vs the free-flight range angle is plotted for various values of $Q_{bo} = r_{bo}V^2_{bo}/\mu$ with $h_{bo} = 10^6$ ft. Note that for values of $Q_{bo} < 1$ there are two values of time of flight for each value of Ψ. The smaller value is for the low trajectory, the larger for the high trajectory. For $Q_{bo} > 1$ the time of flight is a single-valued function of the range since the distinction between high and low trajectory had disappeared.

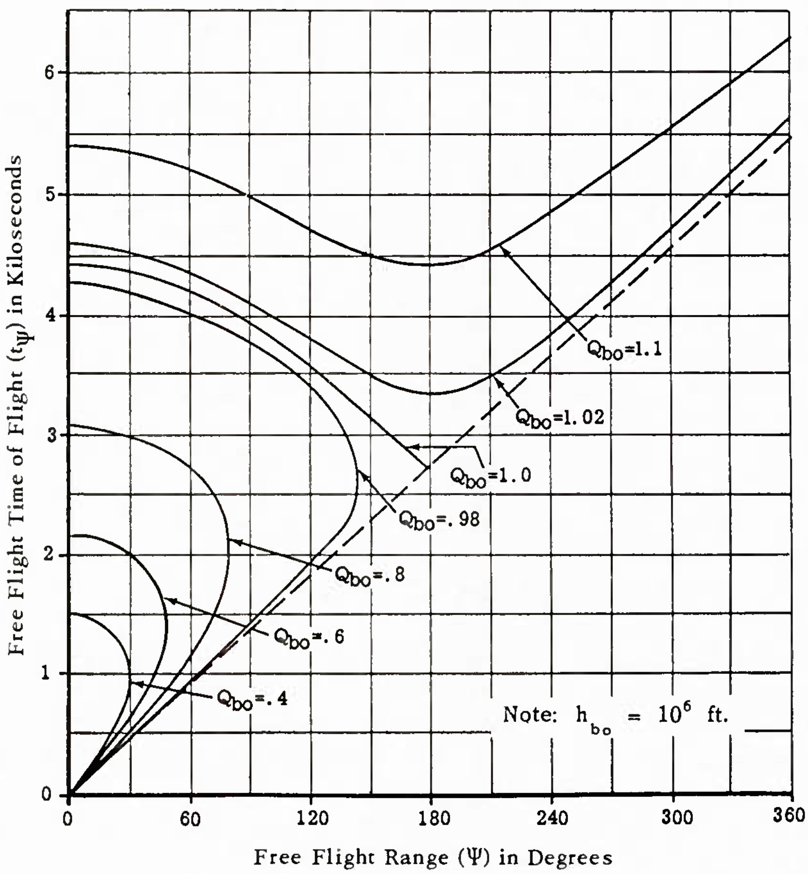


Figure 3. Free flight time of flight.

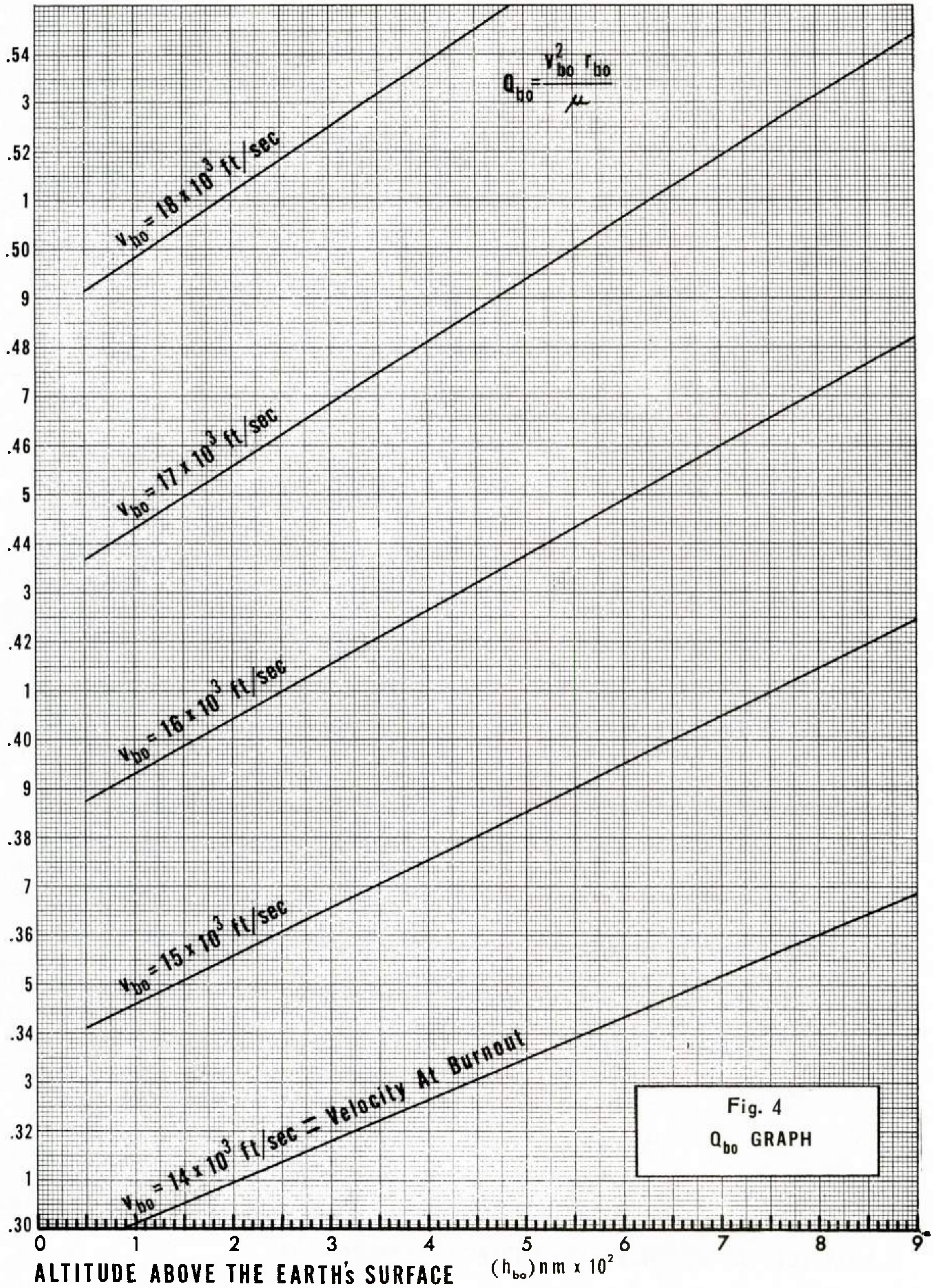


Fig. 4

Q_{bo} GRAPH

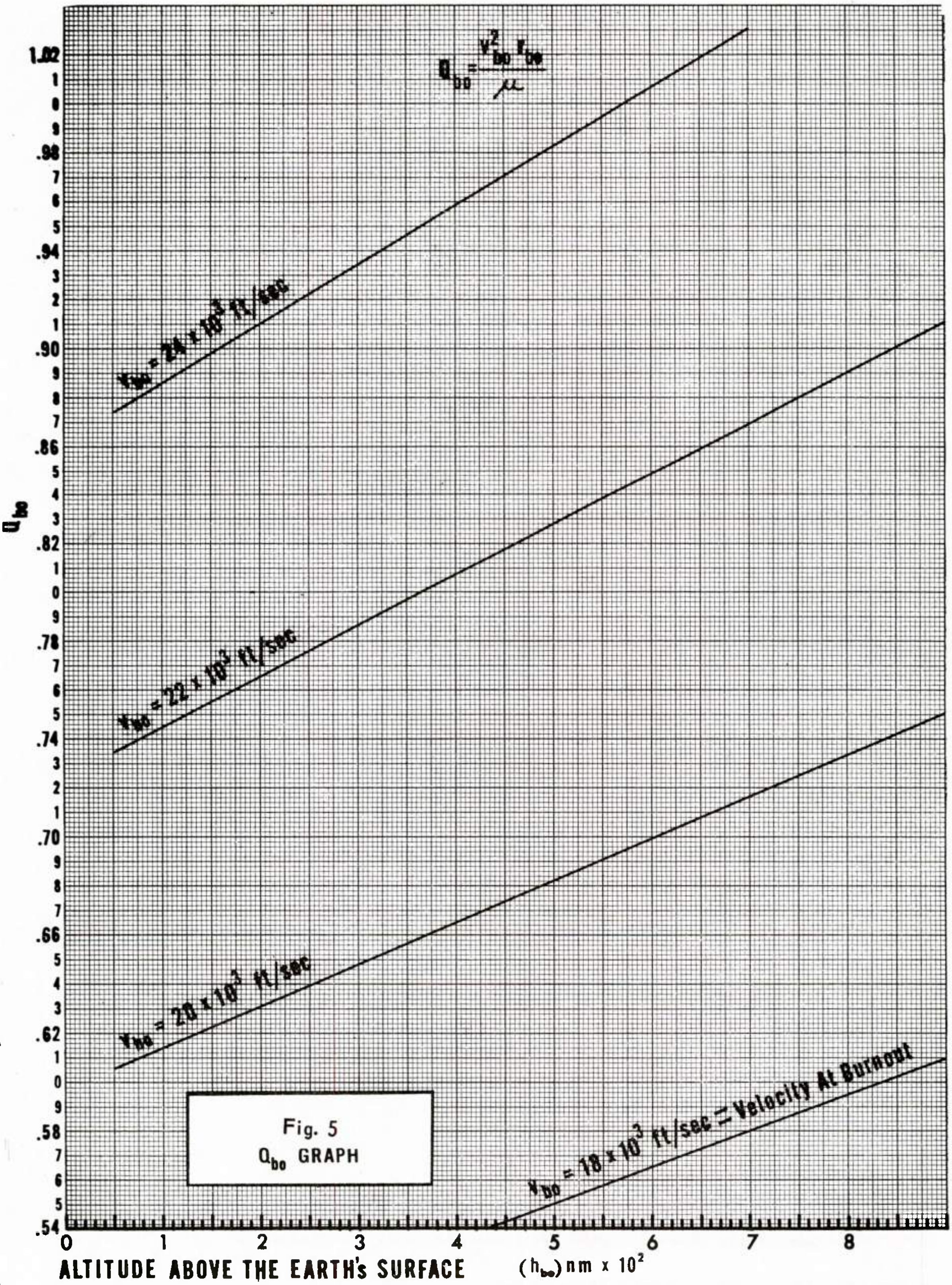


Fig. 5
Q_{bo} GRAPH

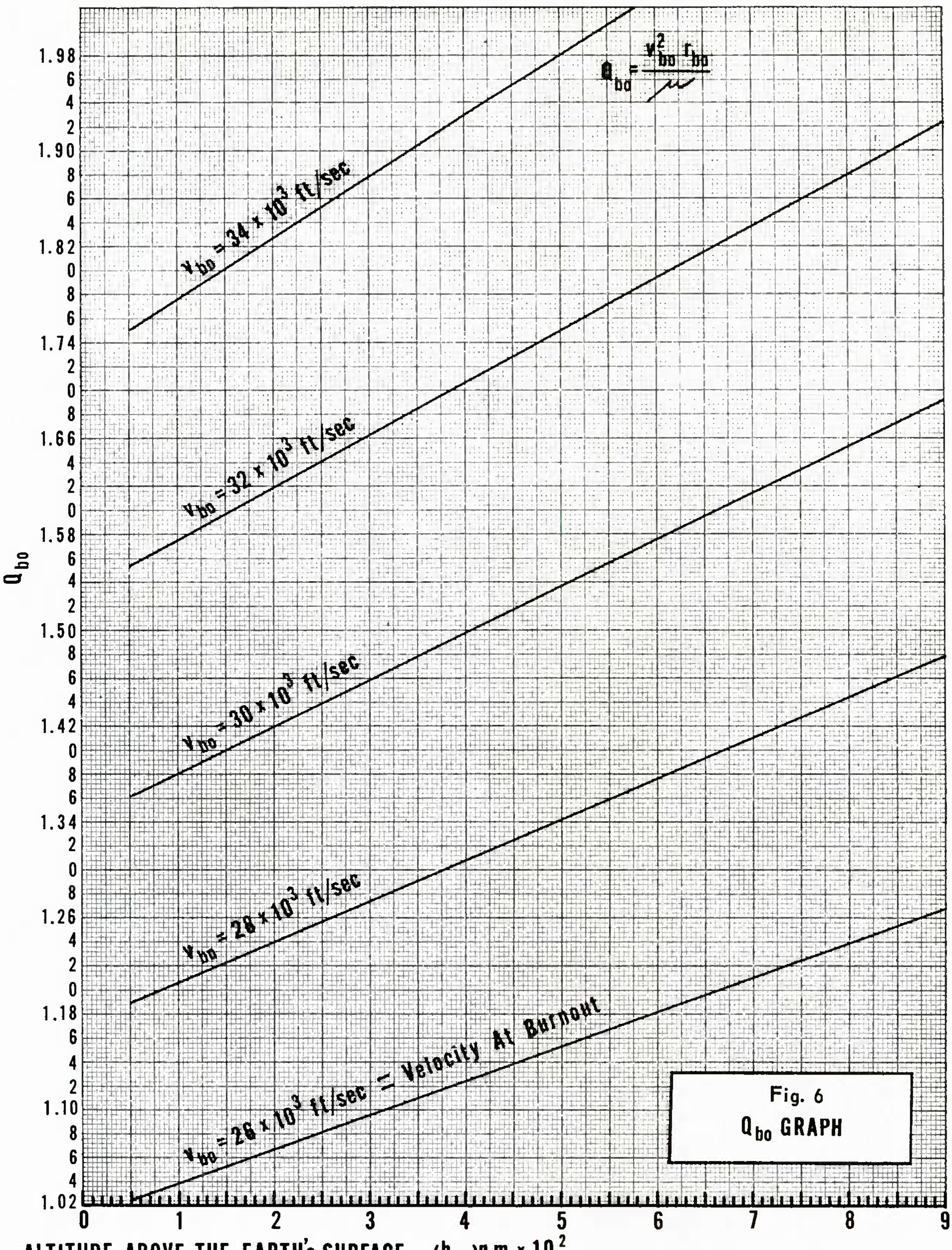


Fig. 6
Q_{bo} GRAPH

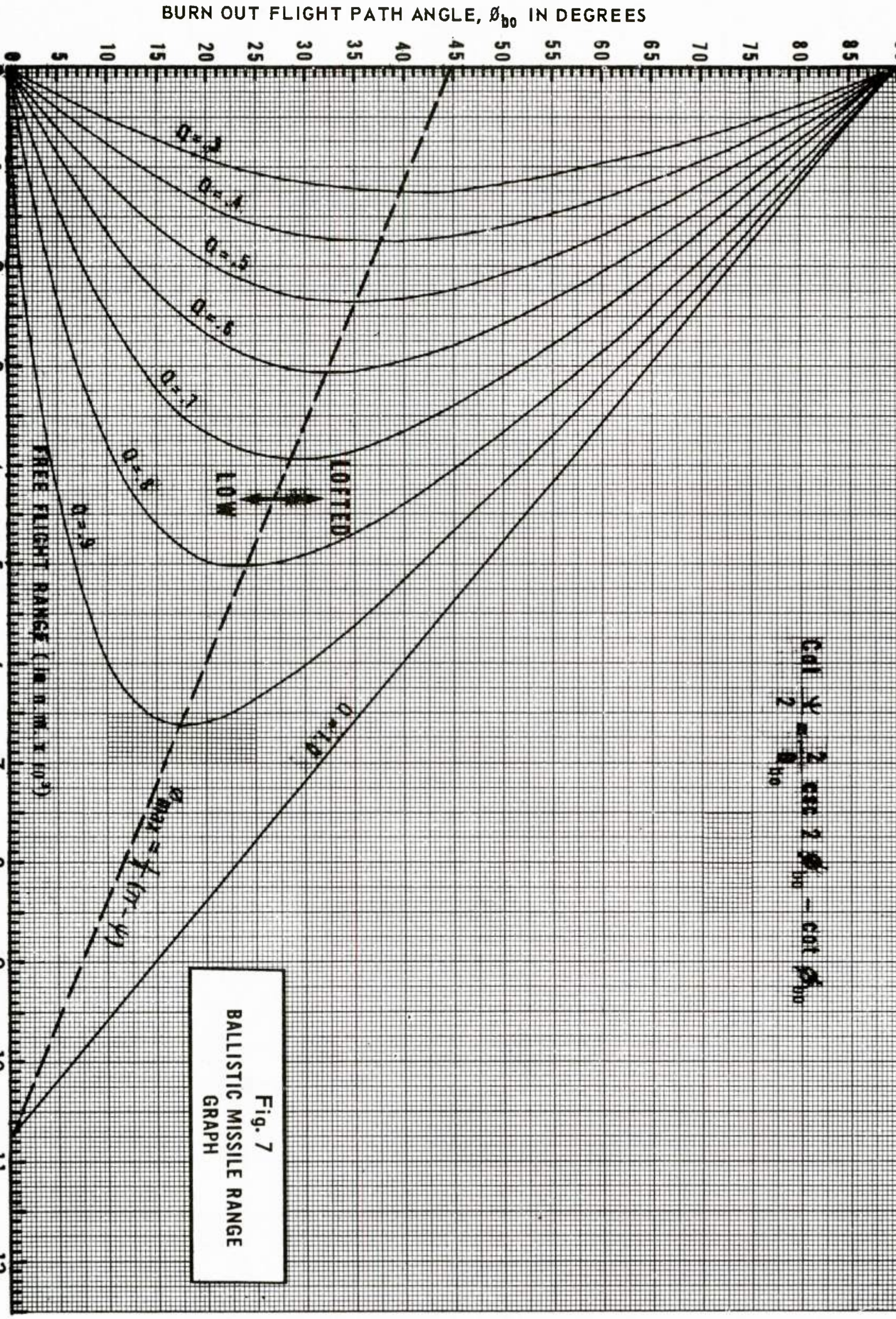


Fig. 7
BALLISTIC MISSILE RANGE GRAPH

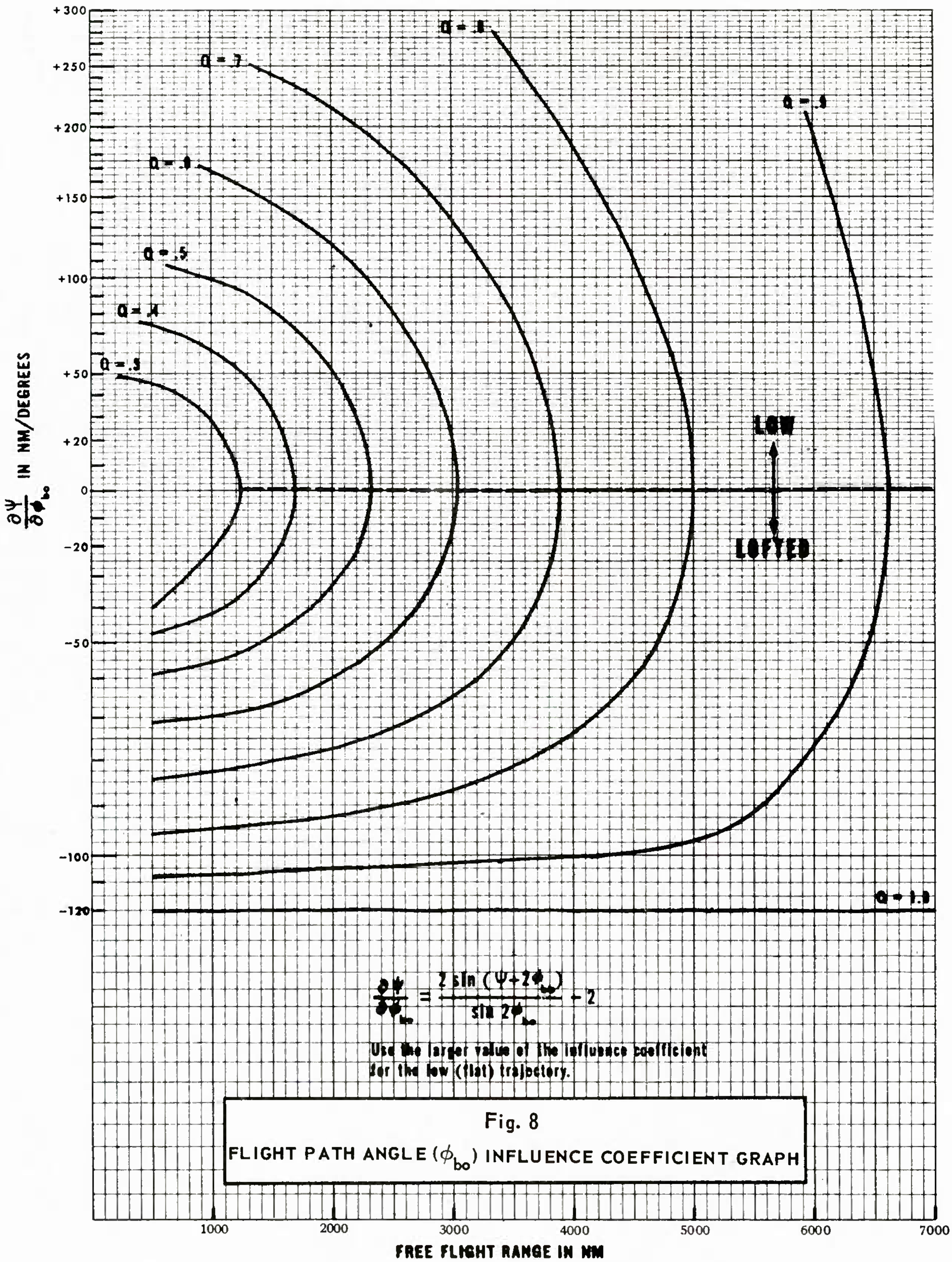


Fig. 8

FLIGHT PATH ANGLE (ϕ_{bo}) INFLUENCE COEFFICIENT GRAPH

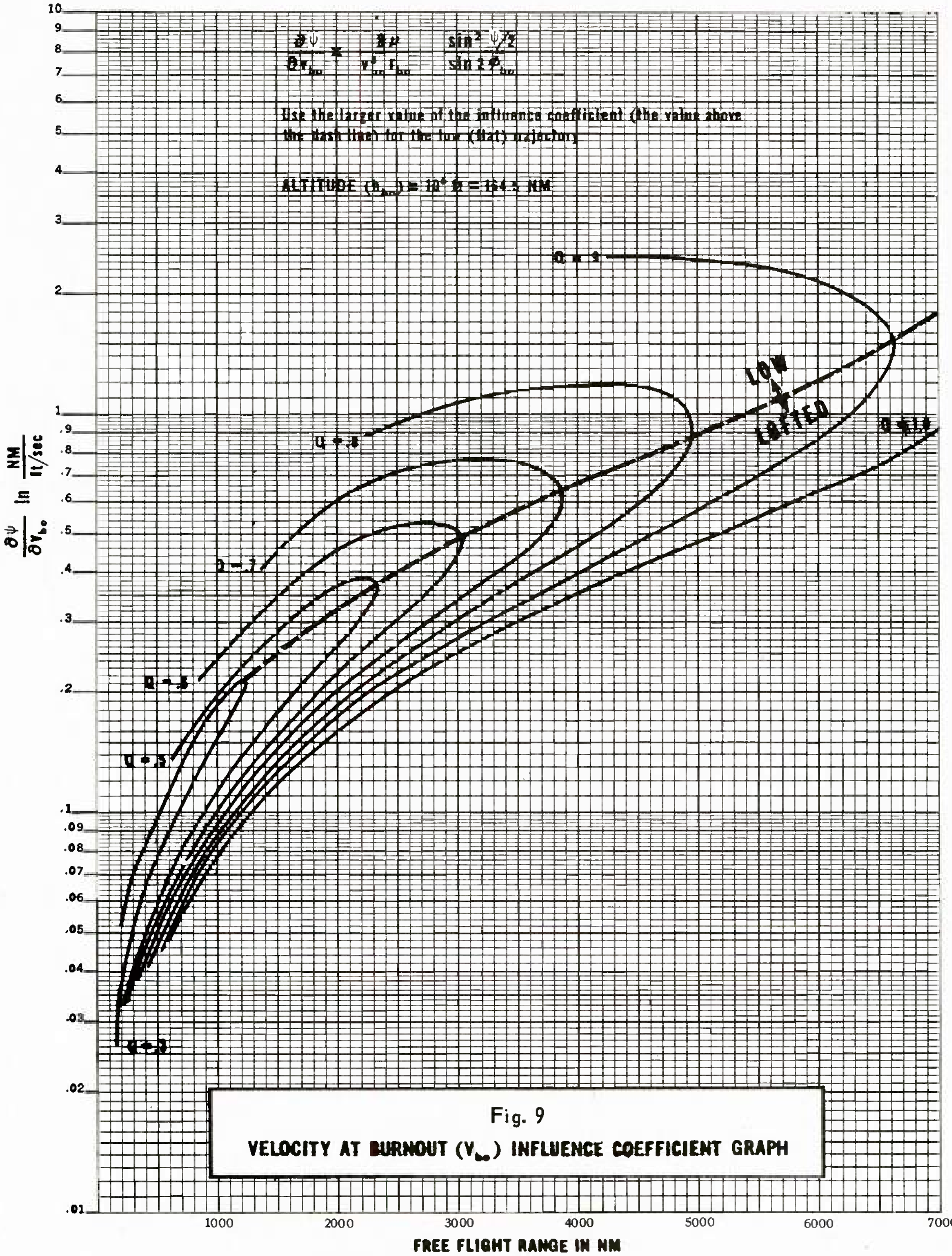


Fig. 9

VELOCITY AT BURNOUT (V_{bo}) INFLUENCE COEFFICIENT GRAPH

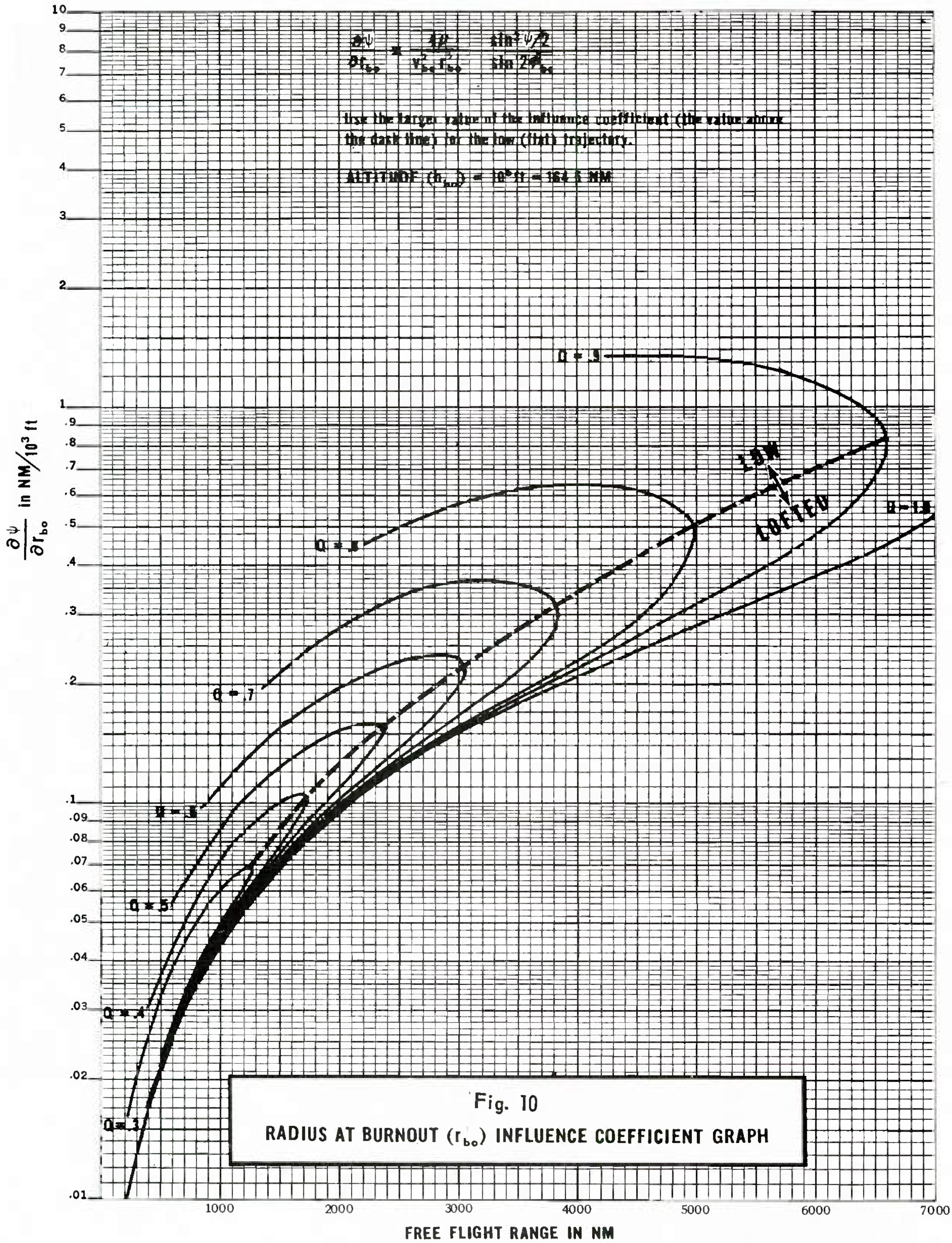


Fig. 10

RADIUS AT BURNOUT (r_{bo}) INFLUENCE COEFFICIENT GRAPH

Appendix D

TIME OF FLIGHT

THE TIME OF FLIGHT of an object traveling on an elliptical path is an important quantity. For example, the time of flight of a ballistic missile must be known in order to compute the apparent motion of the target due to the rotation of the earth. In order to derive an equation for the time of flight of a vehicle on an elliptical path, it is convenient to define three quantities known as anomalies.

In Fig. 1 an elliptical orbit with center of force at a focus F is shown. A circle of diameter equal to the major axis of the ellipse is drawn with center at the center of the ellipse, O. The anomalies of the general point R on the ellipse are defined as follows:

True Anomaly, ν—The angle BFR, measured from periapsis to the designated point on the flight path.

Eccentric Anomaly, u—A perpendicular to the major axis is dropped from R. This line intersects the circle at point S. The eccentric anomaly is the angle BOS. In the equations of this appendix, u is measured in radians.

Mean Anomaly, M—If t is the time of flight from the periapsis B to point R, and P is the period, the mean anomaly is:

$$M = 2\pi \frac{t}{P} \tag{1}$$

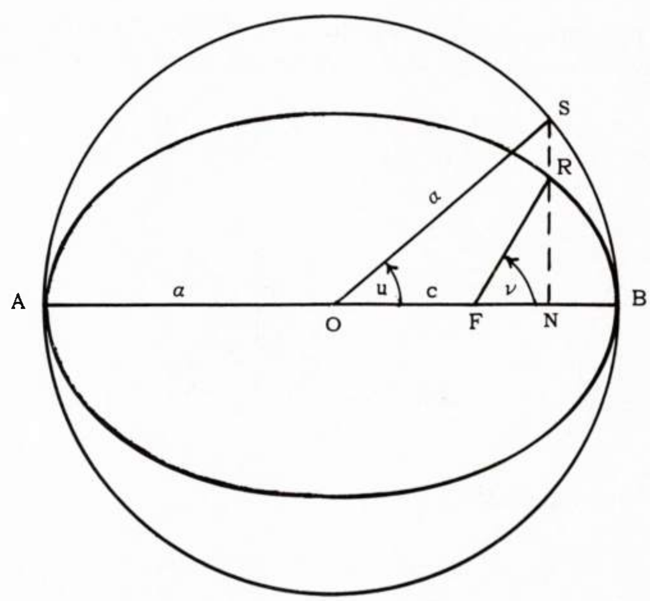


Figure 1. Geometry of the anomalies.

Note: All three anomalies are zero at the periapsis; all three are π at the apoapsis; if one is in the range O to π, the other two are also; and, if one is in the range π to 2π, the other two are also.

To obtain M, and hence t, apply Kepler's second law; area is swept out at a constant rate.

For the general point R, the area swept following periapsis passage is sector FBR. Then:

$$\frac{t}{P} = \frac{\text{Area FBR}}{\text{Area of ellipse}} \tag{2}$$

To determine these areas, write the equation of the ellipse and the auxiliary circle in rectangular coordinates, and compare the ordinates (y values) of the general point.

For the ellipse:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \text{ or } y_{\text{ellipse}} = \frac{b}{a}\sqrt{a^2 - x^2}$$

For the auxiliary circle of radius a:

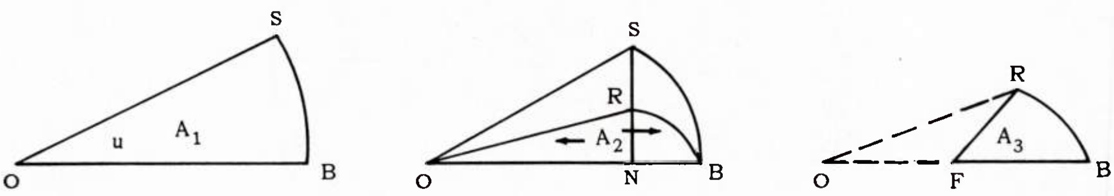
$$\frac{x^2}{a^2} + \frac{y^2}{a^2} = 1 \text{ or } y_{\text{circle}} = \sqrt{a^2 - x^2}$$

or

$$\frac{y_{\text{ellipse}}}{y_{\text{circle}}} = \frac{b}{a} \tag{3}$$

Equation (3) is important in subsequent comparisons of lengths and areas. For example, the area of the ellipse and the auxiliary circle in Figure 1 are related by:

$$\frac{\text{Area (ellipse)}}{\text{Area (circle)}} = \frac{\pi ab}{\pi a^2} = \frac{b}{a}$$



Examine the following areas:

In Figure 2(a) the area of the circular sector is the fraction $\left(\frac{\text{u radians}}{2\pi \text{ radians}}\right)$ of the area (π a^2) of the auxiliary circle; therefore:

$$A_1 = \frac{u}{2\pi}(\pi a^2) = \frac{u a^2}{2}$$

In Figure 2(b) triangles ONR and ONS have the same base, and their heights (therefore areas) are in the ratio $\frac{NR}{NS} = \frac{b}{a}$. The areas of the elliptical segment NBR and circular segment NBS are in the same ratio. Therefore:

$$A_2 = \frac{b}{a} A_1 = \frac{b}{a}\left(\frac{u\,a^2}{2}\right) = \frac{u\,ab}{2}$$

The area (A_3) swept following periapsis passage is obtained by subtracting the triangle OFR from A_2. The base of the triangle OFR is $c = a\epsilon$, and, in terms of u, its height is $\frac{b}{a}(a \sin u) = b \sin u$.

Then:

$$A_3 = A_2 - \tfrac{1}{2}(a\epsilon)(b \sin u) = \frac{u\,ab}{2} - \frac{\epsilon\,ab}{2} \sin u = \frac{ab}{2}(u - \epsilon \sin u)$$

From Equation (2):

$$\frac{t}{P} = \frac{A_3}{\pi\,ab} = \frac{\frac{ab}{2}(u - \epsilon \sin u)}{\pi\,ab} = \frac{1}{2\pi}(u - \epsilon \sin u)$$

$$\text{or } t = \frac{P}{2\pi}(u - \epsilon \sin u)$$

$$\text{Where } P = 2\pi\sqrt{\frac{a^3}{\mu}} \qquad (4)$$

From Equation (1)

$$M = 2\pi \frac{t}{P} = u - \epsilon \sin u \qquad (5)$$

Then the time from periapsis to any point on an elliptical orbit can be found from the definition of the mean anomaly and the known period.

$$t = \frac{P}{2\pi}(u - \epsilon \sin u) = \sqrt{\frac{a^3}{\mu}}(u - \epsilon \sin u) \qquad (6)$$

The time of flight between any two points in general (1 and 2) depends upon the difference in the mean anomalies of the points:

$$t_{1\to 2} = \frac{P}{2\pi}(M_2 - M_1) = \sqrt{\frac{a^3}{\mu}}(M_2 - M_1)$$

$$t_{1\to 2} = \sqrt{\frac{a^3}{\mu}}\left[(u_2 - \epsilon \sin u_2) - (u_1 - \epsilon \sin u_1)\right] \qquad (7)$$

Note: If u_1 is greater than u_2, periapsis is passed in the transit time, and Equation (6) results in a negative value. This value is the difference between the correct time of flight and one period; to eliminate the problem of negative values, replace u_2 with $2\pi + u_2$.

For most practical problems it is necessary to relate the eccentric anomaly u to the true anomaly ν.
From Figure 1:

$$\cos u = \frac{c + r\cos\nu}{a}$$

$$\text{Since } c = a\epsilon \text{ and } r = \frac{a(1-\epsilon^2)}{1+\epsilon\cos\nu}$$

$$\cos u = \frac{a\epsilon + \dfrac{a(1-\epsilon^2)\cos\nu}{1+\epsilon\cos\nu}}{a}$$

$$\cos u = \frac{\epsilon(1+\epsilon\cos\nu) + (1-\epsilon^2)\cos\nu}{1+\epsilon\cos\nu}$$

$$\cos u = \frac{\epsilon + \epsilon^2\cos\nu + \cos\nu - \epsilon^2\cos\nu}{1+\epsilon\cos\nu}$$

$$\cos u = \frac{\epsilon + \cos\nu}{1+\epsilon\cos\nu} \tag{8}$$

Similarly:

$$\sin u = \frac{\sqrt{1-\epsilon^2}\sin\nu}{1+\epsilon\cos\nu} \tag{9}$$

The path of a ballistic missile (Fig. 3) is nearly symmetrical from burnout to reentry.

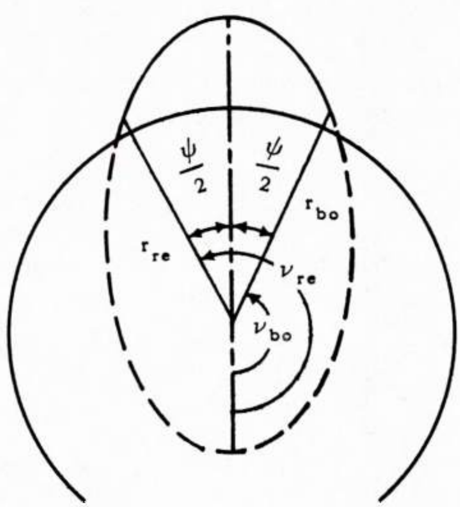


Figure 3. Geometry of ballistic missile time of flight.

If the range angle is ψ, the following relationships are noted:

$$\nu_{bo} = \pi - \frac{\psi}{2} \qquad \cos \nu_{bo} = -\cos \frac{\psi}{2} \qquad \sin \nu_{bo} = \sin \frac{\psi}{2}$$

$$\nu_{re} = \pi + \frac{\psi}{2} \qquad \cos \nu_{re} = -\cos \frac{\psi}{2} \qquad \sin \nu_{re} = -\sin \frac{\psi}{2}$$

$$\nu_{re} = 2\pi - \nu_{bo} \qquad \cos \nu_{re} = \cos \nu_{bo} \qquad \sin \nu_{re} = -\sin \nu_{bo}$$

Assuming 1 to be the burnout point and 2 to be the reentry point, Equation (7) becomes:

$$t_{bo \rightarrow re} = t_{\psi} = \sqrt{\frac{a^3}{\mu}} \left[(2\pi - u_{bo} + \epsilon \sin u_{bo}) - (u_{bo} - \epsilon \sin u_{bo})\right]$$

$$t_{\psi} = \sqrt{\frac{a^3}{\mu}} (2\pi - 2u_{bo} + 2\,\epsilon \sin u_{bo})$$

$$t_{\psi} = \sqrt{\frac{a^3}{\mu}} (\pi - u_{bo} + \epsilon \sin u_{bo}) \tag{10}$$

From Equation (8):

$$\cos u_{ob} = \frac{\epsilon + \cos \nu_{bo}}{1 + \epsilon \cos \nu_{bo}} = \frac{\epsilon - \cos \frac{\psi}{2}}{1 - \epsilon \cos \frac{\psi}{2}} \tag{11}$$

Appendix E

PROPULSION

Discussion of Vehicle Net Acceleration at Launch

THE ACCELERATION of a launch vehicle as it leaves the pad is caused by the difference between the engine thrust and the vehicle weight. From Newton's second law:

$$f = Ma$$

or

$$a = \frac{f}{M}$$

The net force (f) in this case is vehicle thrust (F) minus vehicle weight (W) at lift-off, and M is the mass of the object acted upon by the force differential, i.e., the vehicle itself. Then:

$$a = \frac{F-W}{M}$$

$$\text{but } M = \frac{W}{g}$$

$$\therefore a = \left(\frac{F-W}{W}\right) g$$

or

$$a = \left(\frac{F}{W} - \frac{W}{W}\right) g = (\Psi - 1)\ g$$

where Ψ is the thrust-to-weight ratio.

$\therefore a = (\Psi - 1)$ g's (Equation 6, Chapter 3)

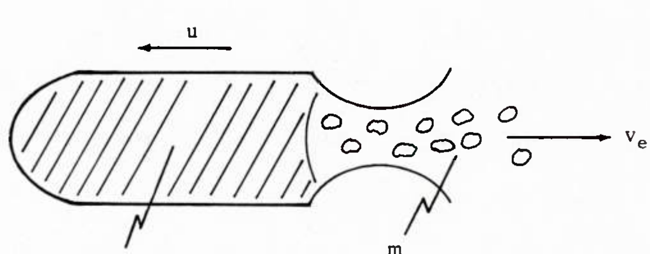


Figure 1

Discussion of Rocket Engine Thrust Equation

The efflux of hot gas from a rocket can be regarded as the ejection of small masses such as m_o (gas molecules) at a high relative velocity v_e with respect to the vehicle, which has mass m and is moving at velocity u. According to Newton's second law, the sum of the unbalanced forces acting on the vehicle is equal to the time rate of change of the vehicle's momentum. The system being considered is the overall vehicle. With these ideas in mind:

$$\Sigma f = \frac{d}{dt}(mv) = m\frac{dv}{dt} + v\frac{dm}{dt}$$

The m is the instantaneous vehicle mass. $\frac{dv}{dt}$ applies to the vehicle and should be $\frac{du}{dt}$. v in the second term applies to the ejected masses which is v_e. And, $\frac{dm}{dt}$ applies to either vehicle or gases since $-\frac{dm}{dt} = \frac{dm_o}{dt}$. Using $-\frac{dm}{dt}$ yields:

$$\Sigma f = m\frac{du}{dt} - v_e\frac{dm}{dt}$$

If the vehicle is operating in a "weightless" environment free of an atmosphere and under steady state conditions, the sum of the unbalanced forces equals 0. Thus:

$$\Sigma f = 0 = m\frac{du}{dt} - v_e\frac{dm}{dt}$$

and

$$m\frac{du}{dt} = v_e\frac{dm}{dt}$$

The left side of the equation represents the constant (steady state) unbalanced force (thrust) on the vehicle ($F = m\frac{du}{dt}$ from Newton's second law). Then:

$$\text{Thrust} = F = v_e\frac{dm}{dt}$$

Since F is constant for steady state, the variables may be separated and integration applied, or one may recognize $\frac{dm}{dt}$ as the mass rate of flow of the rocket exhaust (the propellants). Either way, for an atmosphere-free environment, momentum thrust is:

$$F = \dot{M}v_e$$

or

$$F = \frac{\dot{W}}{g}v_e$$

When the vehicle flies through an ambient fluid such as the earth's atmosphere, the flow around it is affected, and the fluid can interact with the rocket exhaust. Under these conditions the thrust relationships must be corrected by the effect of the pressure forces which act on the surfaces of the vehicle body. The figure below shows the ambient pressure acting uniformly on the outer surface of a rocket chamber and the gas pressures on the inner surface.

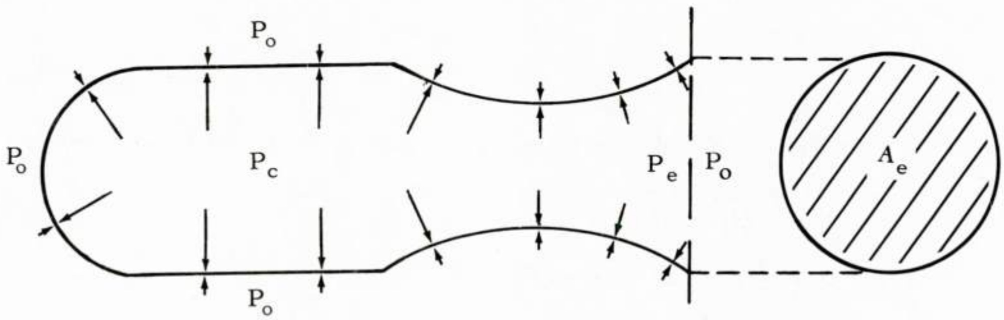


Figure 2

The size of the arrows indicates the relative magnitude of the pressure. The axial thrust can be determined by integrating all the pressure acting on areas which can be projected on a plane normal to the nozzle axis. The forces acting radially outward are appreciable but balance each other and do not contribute to axial thrust.

If the plane upon which all integrated pressures are projected is that of the engine exhaust, one finds that at the exit area A_e there is an imbalance between ambient (P_o) and the local pressure of the exhaust (P_e), the difference being ($P_e - P_o$). The differential force involved is evaluated over the area of the nozzle exit and is equal to $A_e(P_e - P_o)$.

The total thrust acting on the rocket then, must be the sum of the momentum and pressure thrusts or:

$$F = \frac{\dot{W}}{g} v_e + A_e (P_e - P_o) \text{ [Equation 1, Chapter 3]}$$

The second term on the right side of this equation (pressure thrust) affects thrust in this way. If the exhaust pressure is less than ambient pressure, the pressure thrust is negative. Because this condition gives low thrust, the nozzle is usually designed so that P_e equals or slightly exceeds P_o. When $P_e = P_o$ maximum thrust for a given propellant and chamber is attained.

Discussion of Ideal Velocity of a Rocket at Thrust Termination

The equation for the magnitude of the "ideal" vehicle velocity at thrust termination is $\Delta v_i = I_{sp}\, g \ln \left(\frac{W_1}{W_2}\right)$. It can be derived from Newton's second law using the following assumptions:

1. Propellant used at a constant rate.

2. Thrust constant.

Newton's second law may be stated in *any consistent units.*

$$F = Ma = f \text{ (no gravity; no drag)} \tag{A}$$

or

$$F = M \frac{dv}{dt} \tag{B}$$

Since propellant is used at a constant rate the mass of the vehicle at any time is:

$$M = M_1 - \dot{M}t \tag{C}$$

Substitute equation (C) in equation (B):

$$F = (M_1 - \dot{M}t) \frac{dv}{dt} \tag{D}$$

Separate variables and integrate:

$$\int_{v_1}^{v_2} dv = \int_{o}^{t_2} \frac{Fdt}{(M_1 - \dot{M}t)} \tag{E}$$

Since thrust is constant:

$$v_2 - v_1 = F \int_{t_i}^{t_2} \frac{dt}{(M_1 - \dot{M}t)} \tag{F}$$

Since propellant is used at a constant rate:

$$v_2 - v_1 = - \frac{F}{\dot{M}} \Big[1n\ (M_1 - \dot{M}t) \Big]_o^{t_2} \tag{G}$$

When $t_i = o$:

$$M_1 - \dot{M}\,(o) = M_1 \tag{H}$$

When $t = t_2$:

$$M_1 - \dot{M}t_2 = M_2 \tag{I}$$

Substitute equations (H) and (I) in equation (G):

$$v_2 - v_1 = - \left[\frac{F}{\dot{M}} \right] \Big[1n\ M_2 - 1n\ M_1 \Big] \tag{J}$$

Note: $-(\ln M_2 - \ln M_1) = (\ln M_1 - \ln M_2)$

$$v_2 - v_1 = \Delta v = \frac{F}{\dot{M}} \ln \left(\frac{M_1}{M_2}\right) \qquad (K)$$

If we use the English Engineering System, mass is measured in slugs (M) and weight is measured in pounds (W).

$$\frac{M_1}{M_2} = \frac{M_1 g}{M_2 g} = \frac{W_1}{W_2} \qquad (L)$$

$$\dot{M} = \frac{\dot{W}}{g} \quad \textit{Converting mass to weight at earth's surface.} \qquad (M)$$

$$I_{sp} = \frac{F}{\dot{W}} \qquad (N)$$

Substitute equations (L) (M) and (N) in equation (K) and obtain:

$$\Delta v_i = I_{sp}\, g \ln \left(\frac{W_1}{W_2}\right) \quad \text{(Equation 7, Chapter 3)} \qquad (O)$$

Discussion of Mass Ratio for Multi-Stage Rockets

Equation (O) from the previous section can now be applied to a multistage launch vehicle if it is assumed that the velocity of each stage has the same direction. Then the magnitude of the vehicle velocity at thrust termination of the last stage is the sum of the velocity magnitudes of each stage. Consider a three stage vehicle with *each stage having the same specific impulse.*

Stage 1: $\Delta v_1 = I_{sp}\ g\ \ln \left(\frac{W_1}{W_2}\right)_1$

Stage 2: $\Delta v_2 = I_{sp}\ g\ \ln \left(\frac{W_1}{W_2}\right)_2$

Stage 3: $\Delta v_3 = I_{sp}\ g\ \ln \left(\frac{W_1}{W_2}\right)_3$

$$\Delta v_i = \Delta v_1 + \Delta v_2 + \Delta v_3 = I_{sp}\,(g)\left[\ln \left(\frac{W_1}{W_2}\right)_1 + \ln \left(\frac{W_1}{W_2}\right)_2 + \ln \left(\frac{W_1}{W_2}\right)_3\right]$$

This equation contains the sum of three natural logarithms in the brackets. The sum of the natural logarithms of several numbers is the logarithm of the product of the numbers. Therefore, the quantity in the brackets is the natural logarithm of the product of the stage mass ratios.

$$\Delta v_i = I_{sp}\, g \ln \left(\frac{W_1}{W_2}\right)_1 \left(\frac{W_1}{W_2}\right)_2 \left(\frac{W_1}{W_2}\right)_3$$

This is the basis for multiplying the individual mass ratios of each stage to get the overall mass ratio of a multistage vehicle.

Discussion of Electric Rocket Thrust Equation

The electric power input and power output of an electric propulsion system can be equated if a parameter called efficiency is placed in the equation. Thus:

(Power in) (efficiency) = power out

In terms of electrical power and kinetic energy per unit time, this equation may be written:

$$\eta p = \frac{\dot{M} v_e^2}{2}$$

where η = efficiency (%)

$\dot{M}$ = mass flow rate of expelled particles $\left(\frac{\text{slugs}}{\text{sec}}\right)$

v_e = exhaust velocity of expelled particles (ft/sec)

p = electric power input (kw)

From the derivation of the rocket thrust equation, momentum thrust (F) is expressed as $\dot{M} v_e$ (there is no appreciable pressure thrust in an electric engine).

$$\text{then } \eta p = \frac{(\dot{M} v_e) v_e}{2} = \frac{F v_e}{2}$$

$$\text{but } F = \dot{M} v_e \text{ or } v_e = \frac{F}{\dot{M}}$$

$$\text{then } \eta p = \frac{(F)\,(F)}{2\dot{M}}$$

$$\text{and } F^2 = 2\eta p\, \dot{M}$$

To make units compatible on both sides of this equation, a conversion constant ($K = 737.56 \frac{\text{ft lb}}{\text{kw sec}}$) must be applied.

This yields:

$$F^2 = 2\, K\, \eta\, p\, \dot{M}$$

$$F = \sqrt{2(737.56)\eta p\, \dot{M}}$$

$$F = 38.4 \sqrt{\eta\, p\, \dot{M}}$$ (Equation 16 at end of Chapter 3)

Discussion of Theoretical Specific Impulse

The principle of conservation of energy may be written for an isentropic process between any two points in a rocket engine. In this discussion the two points under consideration are the combustion chamber and nozzle exit. Begin with the energy equation in which the decrease in specific enthalpy is equal to the increase in kinetic energy of the flowing gases.

$$\therefore h_c - h_e = \frac{1}{2gJ}(v_e^2 - v_c^2)$$

where: h_c = specific enthalpy in combustion chamber
h_e = specific enthalpy at nozzle exit
g = 32.2 ft/sec²
J = mechanical equivalent of heat (778 ft-lb/BTU)
v_e = velocity of exhaust gases
v_c = velocity of gases in the combustion chamber

Since $v_c \approx O$ in the longitudinal direction due to the random motion of gases in the combustion chamber, neglect its contribution and rearrange the equation to give:

$$v_e^2 = 2g\,J\Delta h \qquad \text{where } \Delta h = h_c - h_e$$

Also, in the isentropic flow of a perfect gas:

$$\Delta h = C_p \Delta T$$

where: C_p = specific heat of a gas at constant pressure
$\Delta T = T_c - T_e$
T_c = temperature of gas in combustion chamber
T_e = temperature of exhaust gases

$$\therefore v_e^2 = 2g\,J\,C_p\,(T_c - T_e)$$

and since $C_p = \left(\frac{k}{k-1}\right)\left(\frac{R}{J}\right)$

where: k = ratio of specific heats

$$R = \frac{R'}{m} = \frac{\text{universal gas constant}}{\text{molecular weight of combustion products}}$$

$$\therefore v_e^2 = 2gR'\left(\frac{k}{k-1}\right)\frac{T_c}{m}\left(1 - \frac{T_e}{T_c}\right)$$

In the isentropic process of a perfect gas:

$$\frac{T_e}{T_c} = \left(\frac{P_e}{P_c}\right)^{\frac{k-1}{k}}$$

$$\therefore v_e^2 = 2g\frac{k}{k-1}R'\frac{T_c}{m}\left[1 - \left(\frac{P_e}{P_c}\right)^{\frac{k-1}{k}}\right]$$

and since R' = 1544 ft-lb/°R mole:

$$v_e^2 = 2\,(32.2)(1544)\ \frac{k}{k-1}\ \frac{T_c}{m}\left[1 - \left(\frac{P_e}{P_c}\right)^{\frac{k-1}{k}}\right]$$

$$\text{or } v_e = 315.1\sqrt{\frac{k}{k-1}\ \frac{T_c}{m}\left[1 - \left(\frac{P_e}{P_c}\right)^{\frac{k-1}{k}}\right]}$$

and since:

$$F = \frac{\dot{W}}{g}\, v_e \text{ (optimum expansion)}$$

$$\frac{F}{\dot{W}} = I_{sp} = \frac{v_e}{g}$$

$$\therefore I_{sp} = \frac{315.1}{32.2}\sqrt{\frac{k}{k-1}\ \frac{T_c}{m}\left[1 - \left(\frac{P_e}{P_c}\right)^{\frac{k-1}{k}}\right]}$$

$$I_{sp} = 9.797\sqrt{\frac{k}{k-1}\ \frac{T_c}{m}\left[1 - \left(\frac{P_e}{P_c}\right)^{\frac{k-1}{k}}\right]} \quad \text{(Equation 9 in Chapter 3)}$$

Discussion of Velocity Variations Contributing to Actual Vehicle Velocity

In the equation:

$v_{ACT} = \Delta v(\text{Ideal}) + v_r - \Delta v_{loss}$ (Equation 8, Chapter 3)

v_{ACT} = magnitude of vehicle velocity at last stage thrust termination

v_r = component of earth's tangential velocity in the direction of the launch azimuth

Δv_{loss} = magnitude of velocity loss due to drag and gravity

The contribution v_r can be calculated from launch azimuth and latitude by the equation:

$$v_r = r_e \omega_e \cos L \sin \beta$$

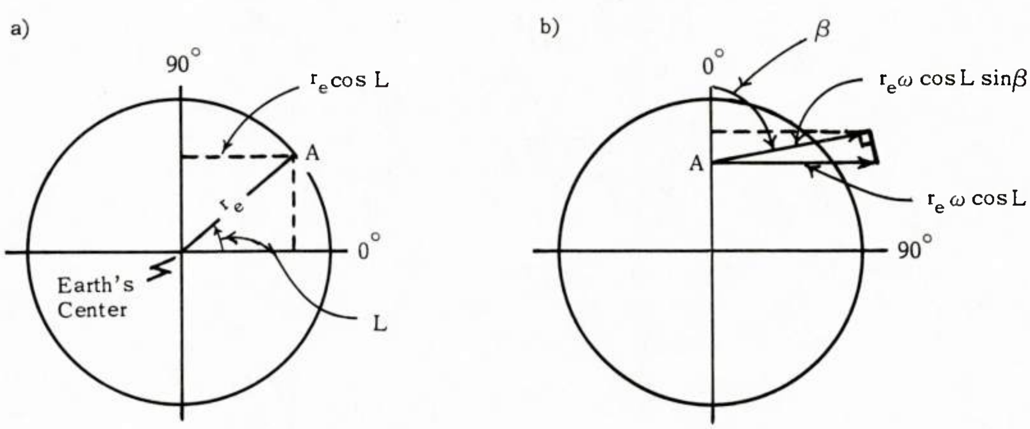


Figure 3

Where: r_e = earth radius (20.9×10^6 feet)

ω_e = earth angular rotation (7.27×10^{-5} radians/sec)

L = launch latitude

β = launch azimuth

In figure 3 a) the effect of launch latitude is illustrated.

In figure 3 b) the effect of launch azimuth is illustrated.
Therefore:

$$v_r = (20.9 \times 10^6)(7.27 \times 10^{-5}) \cos L \sin \beta$$

$$v_r = 1520 \cos L \sin \beta$$

Notice that the maximum contribution to the actual velocity by the effect of the earth's rotation is for an easterly launch at the equator ($L = 0°$) ($\beta = 90°$).

The term Δv_{loss} is calculated from:

$$\Delta v_{loss} = \int g \sin \phi \, dt + \int \frac{\rho v^2 C_d A \, dt}{2M}$$

Where: g = gravitational force (ft/sec²)

ϕ = angle between flight path and perpendicular to radius vector at any altitude (degrees)

dt = time increment (sec).

ρ = atmospheric density (slugs/ft³)

v = vehicle velocity (ft/sec)

C_d = coefficient of drag

A = effective vehicle area (ft²)

M = vehicle mass (slugs)

Calculation of Δv_{loss} is a complex iterative process. A good approximation of Δv_{loss} is 5000 ft/sec for present launch vehicles.
Data for an actual Titan II launch provides a good example of Δv_{loss}:

$$v_r = 1100 \text{ ft/sec}$$

$$\Delta v_1 = 12{,}500 \text{ ft/sec}$$

$$\Delta v_2 = 16{,}900 \text{ ft/sec}$$

$$\Delta v \text{ (Ideal)} = 29{,}400 \text{ ft/sec}$$

$$\text{but } \Delta v_{ACT} = 25{,}756 \text{ ft/sec}$$

$$\therefore \Delta v_{loss} = 4{,}744 \text{ ft/sec}$$

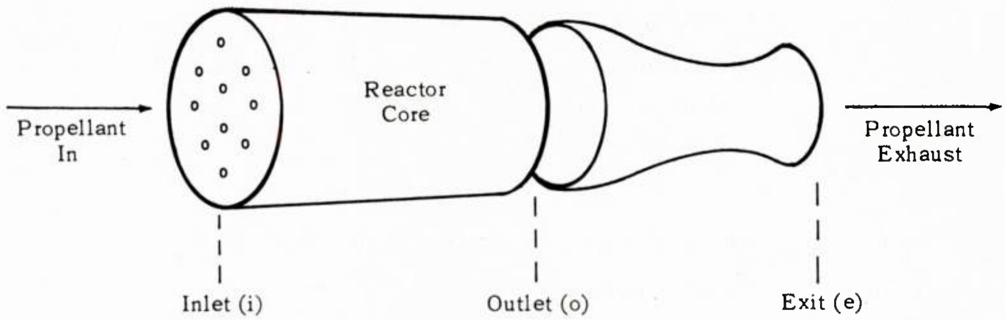


Figure 4

Discussion of Nuclear Rocket Thrust Equation

In the discussion of theoretical specific impulse which occurs earlier in this appendix, the decrease in enthalpy of the exhaust gases flowing through the nozzle of a chemical rocket engine was set equal to the increase in kinetic energy of the flowing gases. This relationship is true also for the propellant flow through the nozzle of a nuclear rocket such as the NERVA engine as shown below.

$$h_o - h_e = \frac{1}{2gJ}(v_e^2 - v_o^2) \qquad \text{(A)}$$

where: h = specific enthalpy (BTU/lb)

g = 32.2 ft/sec^2

v = velocity (ft/sec)

J = mechanical equivalent of heat (778 $\frac{\text{ft lb}}{\text{BTU}}$)

If Q represents total reactor thermal power (BTU/sec) generation, then $Q/\dot{W}$ = q represents the specific input of heat (BTU/lb) available for transfer to the propellant when $\dot{W}$ is the propellant flow rate (lb/sec).

Because the propellant will not receive all thc reactor heat generated (losses and inefficiency in heat transfer), the change in the propellant specific enthalpy across the core will be:

$$h_o - h_i = \nu q$$

where: ν = fraction of q delivered to the propellant

then: $h_o = \nu q + h_i$ (B)

Substituting h_o from equation (B) into equation (A) yields:

$$\nu q + h_i - h_e = \frac{1}{2gJ}(v_e^2 - v_o^2) \qquad \text{(C)}$$

v_o^2 is very small with respect to v_e^2 and may be neglected. And the heat addition and expansion processes occur at nearly a constant pressure so that:

$$h_i - h_e = \int_e^i C_p \, dT \tag{D}$$

where: C_p = propellant constant pressure specific heat $\left(\frac{BTU}{lb\,°R}\right)$

T = propellant bulk (average) temperature (°R)

It appears that the propellant to be used in nuclear rockets in the foreseeable future will be hydrogen. The remainder of this discussion will assume H_2 to be the propellant.

For gaseous H_2, C_p varies between 3.4 and 5.0 $\frac{BTU}{lb\,°R}$ over a temperature range of 600 to 5000° R. This range of variation could be overcome by approximating C_p for hydrogen by $C_p = 2.857 + 2.867\,(10^{-4})\,T + 9.92T^{-1/2}$ (See pg 61 of *Thermodynamics* by Faires, listed in the Bibliography at the end of Chapter 3 of this text, for a fuller discussion.) This could be used in equation (D) and integration applied, but the resulting expression is difficult to use.

The ratio of specific heats ($k = C_p/C_v$) for H_2 over the same temperature range varies only between 1.3 and 1.4. Using this parameter makes the resulting equations easier to use.

$$\text{Since } C_p = \frac{k}{k-1}\left(\frac{R}{J}\right)$$

where: R = gas constant $\left(\frac{ft\,lb}{lb\,°R}\right)$

$$h_i - h_e = \int_e^i \frac{k}{k-1}\left(\frac{R}{J}\right) dT$$

If an average of k between 1.3 and 1.4 is used. Then:

$$h_i - h_e = \frac{k}{k-1}\left(\frac{R}{J}\right)\int_e^i dT$$

$$\text{and: } h_i - h_e = -\frac{k}{k-1}\left(\frac{R}{J}\right)(T_e - T_i) \tag{E}$$

If equation (C) is solved for v_e (with $v_o \approx 0$) and $h_i - h_e$ from equation (E) is substituted:

$$v_e = \sqrt{2gJ\left[\nu\, q - \frac{k}{k-1}\left(\frac{R}{J}\right)(T_e - T_i)\right]} \tag{F}$$

Assuming the best case in which exhaust and ambient pressure are equal and pressure thrust equals zero:

$$F = \frac{\dot{W}}{g} v_e = \text{thrust (lb)}$$

Using equation (F) yields:

$$F = \frac{\dot{W}}{g}\sqrt{2\,g\,J\left[\nu\,q - \frac{k}{k-1}\left(\frac{R}{J}\right)(T_e - T_i)\right]} \qquad \text{(G)}$$

$$\text{or: } F = \frac{1}{g}\sqrt{2\,g\,J\,\dot{W}\left[\nu\,\dot{W}\,q - \dot{W}\left(\frac{k}{k-1}\right)\left(\frac{R}{J}\right)(T_e - T_i)\right]}$$

Recall that $\dot{W}\ q = Q$

$$\text{then: } F = \frac{1}{g}\sqrt{2\,g\,J\,\dot{W}\left[\nu Q - \dot{W}\left(\frac{k}{k-1}\right)\left(\frac{R}{J}\right)(T_e - T_i)\right]} \qquad \text{(H)}$$

For these constant values:

$$\begin{aligned} g &= 32.2\ \text{ft/sec}^2 \\ J &= 778.2\ \text{ft lb/BTU} \\ k &= \text{average value of } 1.35 \\ R &= 766.54\ \frac{\text{ft lb}}{\text{lb }^\circ\text{R}}\ \text{(for hydrogen)} \end{aligned}$$

equation (H) becomes:

$$F = 6.94\sqrt{\dot{W}\,[\nu\,Q - 3.76\,\dot{W}\,(T_e - T_i]} \qquad \text{(I)}$$

In equation (I) Q, reactor thermal power, has units of BTU/sec. However, reactor power is usually represented in megawatts (MW). Employing the conversion factor 947 BTU/sec/MW in equation (I) yields:

$$F = 6.94\sqrt{\dot{W}\,[947\,\nu\,Q - 3.76\,\dot{W}\,(T_e - T_i)]}$$

(Equation 15, at end of Chapter 3)

where: F = thrust (lb)

$\dot{W}$ = hydrogen propellant flow rate (lb/sec)

Q = reactor *thermal* power (MW)

T = propellant bulk temperature (°R)

The parameter ν deserves further comment. It involves the efficiency of heat transfer between reactor structure and the propellant which, in turn, involves the study of heat transfer during forced convection. It involves the study of convective heat transfer under laminar and turbulent flow conditions. And, it involves the study of heat loss from the reactor core. These subjects are not within the scope of this handbook.

Appendix F

ACCURACY REQUIREMENTS FOR ORBITAL GUIDANCE

During the three phases of space flight—the injection phase, the midcourse trajectory, and the terminal phase—maintaining accuracy in guidance is a complex problem but one that is important to solve. The analysis given here provides a simple method of approximating the accuracy requirements for guidance during orbital missions. The orbit is assumed to be about a spherical, nonrotating earth in an environment free of atmospheric drag.

Consider first a satellite injected into orbit at a point r_{bo} from the center of the earth with a burnout speed v_{bo}. From Chapter 2, the general equation for an elliptical orbit is as follows:

$$v_{bo} = \sqrt{\frac{2\mu}{r_{bo}} - \frac{\mu}{a}}$$

or

$$a = \frac{\mu\, r_{bo}}{2\mu - v_{bo}^2\, r_{bo}}$$

Consider small errors (δ) in burnout radius δr_{bo} and burnout speed δv_{bo} and their effect δa upon the semimajor axis. Define δr_{bo} and δv_{bo} as small errors in burnout radius and burnout velocity.

$$r_{bo} \gg \delta r_{bo}; \quad v_{bo} \gg \delta v_{bo}$$

The semi-major axis of the ellipse defined by the satellite injected into the orbit, with a flight path angle of zero degrees, a burnout radius $r_{bo} + \delta r_{bo}$, and burnout velocity $v_{bo} + \delta v_{bo}$, is $a + \delta a$.

Then:

$$a + \delta a = \frac{\mu(r_{bo} + \delta r_{bo})}{2\mu - (v_{bo} + \delta v_{bo})^2\,(r_{bo} + \delta r_{bo})}$$

Since (δv_{bo}) is very small:

$$(v_{bo} + \delta v_{bo})^2 \approx v_{bo}^2 + 2 v_{bo}\, \delta v_{bo}$$

$$\therefore a + \delta a = \frac{\mu(r_{bo} + \delta\, r_{bo})}{2\mu - v_{bo}^2\, r_{bo} - 2 v_{bo}\, r_{bo}\, \delta v_{bo} - v_{bo}^2 \delta r_{bo}}$$

Subtracting:

$$\delta a = \frac{\mu r_{bo} + \mu \delta r_{bo}}{2\mu - v_{bo}^2 r_{bo} - 2v_{bo} r_{bo} \delta v_{bo} - v_{bo}^2 \delta r_{bo}} - \frac{\mu r_{bo}}{2\mu - v_{bo}^2 r_{bo}}$$

Making a common denominator of:

$$(2\mu - v_{bo}^2 r_{bo} - 2v_{bo} r_{bo} \delta v_{bo} - v_{bo}^2 \delta r_{bo}) \ (2\mu - v_{bo}^2 r_{bo})$$

$$\delta a = \frac{2\mu^2 r_{bo} + 2\mu^2 \delta r_{bo} - \mu v_{bo}^2 r_{bo}^2 - \mu v_{bo}^2 \delta r_{bo} r_{bo}}{\text{Common denominator}}$$

$$- \frac{2\mu^2 r_{bo} + \mu v_{bo}^2 r_{bo}^2 + 2\mu v_{bo} r_{bo}^2 \delta v_{bo} + \mu v_{bo}^2 r_{bo} \delta r_{bo}}{\text{Common denominator}}$$

Combining terms in numerator:

$$\delta a = \frac{2\mu^2 \delta r_{bo} + 2\mu v_{bo} r_{bo}^2 \delta v_{bo}}{(2\mu - v_{bo}^2 r_{bo} - 2v_{bo} r_{bo} \delta v_{bo} - v_{bo}^2 \delta r_{bo}) (2\mu - v^2_{bo} r_{bo})}$$

Since 2μ and $v_{bo}^2 r_{bo}$ are much greater than either

$2v_{bo} r_{bo} \delta v_{bo}$ or $v_{bo}^2 \delta r_{bo}$ then,

$$\delta a = \frac{2\mu^2 \delta r_{bo} + 2\mu v_{bo} r_{bo}^2 \delta v_{bo}}{(2\mu - v_{bo}^2 r_{bo})^2}$$

$$\frac{\delta a}{a} = \left[\frac{2\mu^2 \delta r_{bo} + 2\mu v_{bo} r_{bo}^2 \delta v_{bo}}{(2\mu - v_{bo}^2 r_{bo})^2}\right]\left[\frac{2\mu - v_{bo}^2 r_{bo}}{\mu r_{bo}}\right]$$

$$= \frac{2\mu \delta r_{bo} + 2v_{bo} r_{bo}^2 \delta v_{bo}}{(2\mu - v_{bo}^2 r_{bo}) (r_{bo})}$$

Multiply top and bottom of right-hand side of above equation by μr_{bo}:

$$\frac{\delta a}{a} = \frac{\mu r_{bo}(2\mu \delta r_{bo})}{(2\mu - v_{bo}^2 r_{bo}) (\mu r_{bo}^2)} + \frac{\mu r_{bo} (2v_{bo} r_{bo}^2 \delta v_{bo})}{(2\mu - v_{bo}^2 r_{bo}) (\mu r_{bo}^2)}$$

$$\frac{\delta a}{a} = a\left[\frac{2}{r_{bo}^2} \delta r_{bo} + \frac{2v_{bo}}{\mu} \delta v_{bo}\right] \tag{1}$$

This equation expresses the change in semimajor axis of an elliptical orbit for small changes or errors in burnout radius and burnout speed.

Next, consider changes in orbital period δP for errors in burnout speed and burnout radius:

$$P = \frac{2\pi a^{3/2}}{\sqrt{\mu}}$$

and $a = \dfrac{\mu r}{2\mu - v^2 r}$, or for burnout conditions

$$a = \frac{\mu r_{bo}}{2\mu - v_{bo}^2 r_{bo}}$$

Therefore:

$$P = \frac{2\pi}{\sqrt{\mu}}\left(\frac{\mu\, r_{bo}}{2\mu - v_{bo}^2\, r_{bo}}\right)^{3/2}$$

Applying partial differentiation:

$$\frac{\partial P}{\partial v} = \frac{2\pi}{\sqrt{\mu}}\,(\mu r_{bo})^{3/2}\left(\frac{-3}{2}\right)\left(2\mu - v_{bo}^2\, r_{bo}\right)^{-5/2}\left(-2v_{bo}\, r_{bo}\right)$$

$$\text{Since } P = \frac{2\pi}{\sqrt{\mu}}\,\frac{(\mu r_{bo})^{3/2}}{(2\mu - v_{bo}^2 r_{bo})^{3/2}}$$

$$\frac{\partial P}{\partial v} = \frac{3P\, r_{bo}\, v_{bo}}{2\mu - v_{bo}^2\, r_{bo}}$$

Multiply top and bottom of right side by μ:

$$\frac{\partial P}{\partial v} = 3P\left(\frac{\mu r_{bo}}{2\mu - v_{bo}^2\, r_{bo}}\right)\left(\frac{v_{bo}}{\mu}\right)$$

$$\frac{\partial P}{\partial v} = \frac{3Pa\, v_{bo}}{\mu}$$

$$\therefore\ \delta P \approx \frac{3a\ P\ v_{bo}}{\mu}\ \delta\ v_{bo} \tag{2}$$

Derivation of Equation 3 is similar to that of Equation 2.

$$\delta P = \frac{3aP}{r_{bo}^2}\ \delta r_{bo} \tag{3}$$

For a nearly circular orbit:

$$a \approx r_{bo} \text{ and } v_{bo}^2 \approx \frac{\mu}{a}$$

Equation 1 can then be simplified to:

$$\delta a = 2\delta r_{bo} + \frac{2a}{v_{bo}}\ \delta v_{bo} \tag{4}$$

The effect of a small error in burnout flight path angle, $\delta\phi_{bo}$, is somewhat complicated for an elliptical orbit. However, if a circular orbit were desired, then the eccentricity of the orbit caused by the velocity vector not being exactly horizontal (i.e.,

$$\epsilon \approx \delta\phi_{bo} \tag{5}$$

where $\delta\phi_{bo}$ is expressed in radians.

Derivation of Equation 5:

The relation between the eccentricity of an orbit and its total specific mechanical energy and angular momentum was given by equation 6, Chapter 2 as:

$$\epsilon = \sqrt{1 + \frac{2EH^2}{\mu^2}}$$

Substituting

$$E = \frac{v^2}{2} - \frac{\mu}{r} \text{ and } H = vr \cos \phi$$

into the first equation:

$$\epsilon = \left[1 + \frac{2}{\mu^2}\left(\frac{v^2}{2} - \frac{\mu}{r}\right) v^2r^2 \cos^2 \phi\right]^{1/2}$$

$$= \left[1 + \frac{v^4r^2}{\mu^2} (1 - \frac{2\mu}{v^2r}) \cos^2 \phi \right]^{1/2}$$

Making use of the definition from Appendix C:

$$Q_{bo} = \frac{V_{bo}{}^2 \, r_{bo}}{\mu}$$

$$\text{then } \epsilon = \left[1 + Q_{bo}{}^2 \left(1 - \frac{2}{Q_{bo}}\right) \cos^2 \phi_{bo}\right]^{1/2}$$

Specializing to the case of a circular orbit, where $Q_{bo} = 1$

$\epsilon = [1 - \cos^2 \phi_{bo}]^{1/2} = [\sin^2 \phi_{bo}]^{1/2} \sin \phi_{bo}$

For a circular orbit, $\epsilon = \sin \delta \phi_{bo}$

For small errors in flight path angle, $(\delta \phi_{bo} = \sin \delta \phi_{bo})$

$$\epsilon = \delta \phi_{bo} \qquad (5)$$

The next step will be to apply the equations for small guidance errors to representative orbits. Since circular orbits have important practical application today, what representative accuracies are required?

Problem 1. An earth satellite is to be injected into a circular orbit at 300 NM above the earth. Assume that guidance errors cause the vehicle to be injected into orbit at 301 NM altitude with a speed at burnout 1 ft/sec higher than desired.

Find: Actual height at apogee of the resulting ellipse.

Circular speed at 300 NM = 24,900 ft/sec

Using Equation 4:

$$\delta a = 2 \, \delta r_{bo} + 2a \frac{\delta V_{bo}}{V_{bo}}$$

$$\delta r_{bo} = 301 \text{ NM} - 300 \text{ NM} = 1 \text{ NM}$$

$$\delta v_{bo} = 1 \text{ ft/sec}$$

$$a = r_e + 300 \text{ NM} = 3440 \text{ NM} + 300 \text{ NM} = 3740 \text{ NM}$$

$$\delta a = (2)\ (1 \text{ NM}) + \frac{(2)\ (3740 \text{ NM})\ (1 \text{ ft/sec})}{24{,}900 \text{ ft/sec}}$$

$$\delta a = 2 \text{ NM} + .300 \text{ NM} = 2.3 \text{ NM}$$

The distance from perigee to apogee equals twice the semimajor axis of the elliptical orbit (Figure 1).

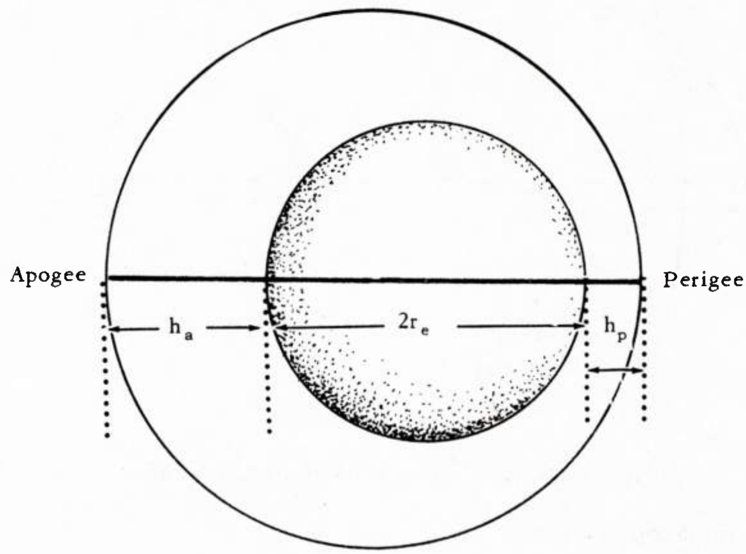


Figure 1. Ellipse formed by orbiting earth satellite.

Therefore,

$$h_a + 2r_e + h_p = 2(a + \delta a)$$

$$h_a + 6880 \text{ NM} + 301 \text{ NM} = 2(3740 + 2.3 \text{ NM})$$

$$h_a = 7484.6 \text{ NM} - 6880 \text{ NM} - 301 \text{ NM} = 303.6 \text{ NM}$$

In the above problem, assume that injection occurred at the exact speed and altitude desired (i.e., $\delta v_{bo} = \delta r_{bo} = 0$), but that the burnout angle was 0.1°. Find the eccentricity of the resulting orbit, and the height of apogee.

$$\epsilon = \delta\phi_{bo} = \frac{0.1°}{57.3°/\text{rad}} = 0.001745$$

$$c = \epsilon\alpha = 0.001745 \times 3740 \text{ NM} = 6.5 \text{ NM}$$

$$r_a = a + c = 3740 + 6.5 = 3746.5 \text{ NM}$$

$$h_a = r_a - r_e = 3746.5 \text{ NM} - 3440 \text{ NM} = 306.5 \text{ NM}$$

Problem 2: Consider next a Hohmann transfer from a 300 NM orbit to 19,360 NM altitude. Find altitude at apogee for a guidance error at perigee of the transfer ellipse of:

(1) δv_{bo} = 1 ft/sec high, no altitude or burnout angle error

(2) δr_{bo} = 6080 ft high, no errors in speed or angle.

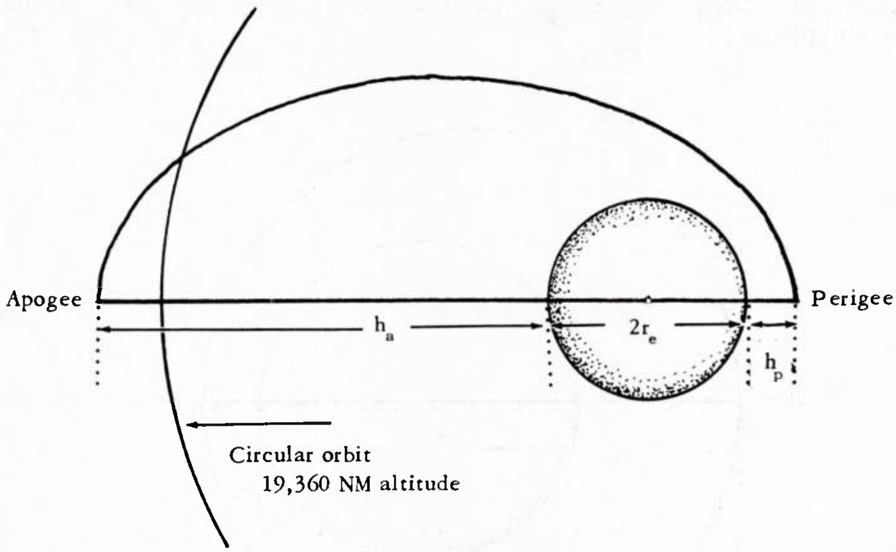


Figure 2. Ellipse formed in the Hohmann transfer.

Using Equation 1:

$$\frac{\delta a}{a} = a\left[\underbrace{\frac{2\,\delta r_{bo}}{r_{bo}^2}}_{\text{Altitude error}} + \underbrace{\frac{2v_{bo}\,\delta v_{bo}}{\mu}}_{\text{Velocity error}}\right]$$

(1) For speed error:

$$\frac{\delta a}{a} = a\,\frac{2v_{bo}\delta v_{bo}}{\mu}$$

$$2a = 19{,}360\text{ NM} + 6880\text{ NM} + 300\text{ NM}$$

$$a = 13{,}270\text{ NM},\ v_{bo} = 32{,}600\text{ ft/sec}$$

$$\delta a = \frac{(13{,}270\text{ NM} \times 6080\text{ ft/NM})^2(2)\ (32{,}600\text{ ft/sec})\ (1\text{ ft/sec})}{14.08 \times 10^{15}\text{ ft}^3/\text{sec}^2}$$

$$= 30{,}200\text{ ft} \approx 5\text{ NM}$$

The distance from perigee to apogee equals twice the semimajor axis of the ellipse attained. See Figure 2 for the major axis.

Therefore,

$$h_a + 2r_e + h_p = 2(a + \delta a)$$

$$h_a + 300\text{ NM} + 6880\text{ NM} = 2(13{,}270\text{ NM} + 5\text{ NM}) = 26{,}550\text{ NM}$$

$$h_a = 19{,}370 \qquad \textit{Answer}$$

(2) For altitude error:

$$\frac{\delta a}{a} = a\left(\frac{2\delta r_{bo}}{r_{bo}^2}\right)$$

$$\delta a = \frac{(13{,}270\ \text{NM})^2(2)\ 1\ \text{NM}}{(3740\ \text{NM})^2} = 25.2\ \text{NM} \approx 25\ \text{NM}$$

$$h_a + h_p + 2r_e = 2(a + \delta a) = 2(13{,}270\ \text{NM} + 25\ \text{NM})$$

$$h_a + 301\ \text{NM} + 6880\ \text{NM} = 26{,}590\ \text{NM}$$

$$h_a = 19{,}409\ \text{NM}$$ *Answer*

Appendix G

ATMOSPHERIC REENTRY

ONLY a few years ago, no one knew whether it was possible to bring a vehicle safely back from space through the atmosphere surrounding the earth. Today atmospheric reentry is almost a routine event. *I*nter-*c*ontinental *b*allistic *m*issile (ICBM) warheads, unmanned data capsules, and manned spacecraft are returned to the surface of the earth to complete their space missions. This achievement within a span of a few years came as the result of experimentation and extensive research into the phenomena of atmospheric reentry.

This chapter examines the reentry phase of a spacecraft in the following areas:

1. The reentry trajectory.
2. Problems caused during reentry.
3. Solutions to the problems.

The chapter gives primary emphasis to reentry into the atmosphere of the earth; however, other planets in the solar system also have atmospheres. Both manned and unmanned missions to Venus and Mars, for example, are currently under study. Since the atmospheres of these planets differ from that of the earth, reentry problems are somewhat different. Essentially, however, the material covered in this chapter is applicable to other atmospheres.

A retrorocket is the only means of slowing a spacecraft from its orbital or interplanetary velocity to a soft landing on the surface of a heavenly body which has no atmosphere. A spacecraft, therefore, must be equipped with a rocket engine that can be fired to slow the vehicle. This retrorocket must be large enough to slow it to the essentially "zero" velocity required for a soft landing. Thus, all U. S. Surveyor spacecraft which made soft landings on the atmosphereless moon had to be equipped with a retrorocket, which was used as shown in Figure 1.

A retrorocket would work just as well on earth. One could have been used to slow the Mercury spacecraft from its 17,000 mph orbital velocity to its soft landing through the earth's atmosphere, thus avoiding some of the reentry problems discussed later in this chapter. But, because the earth is much larger than the moon and the Mercury spacecraft was considerably heavier than Surveyor, an impracticably large retrorocket would have been required. Another, more economical way to slow spacecraft returning to the surface of the earth is to use the aerodynamic drag force developed during reentry into the atmosphere.

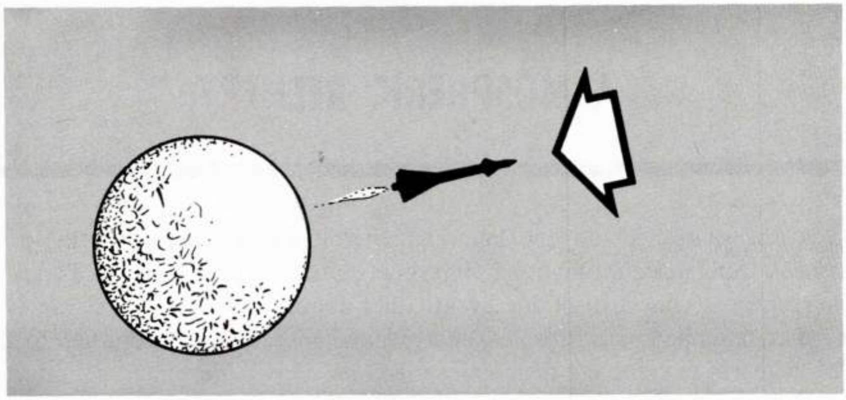


Figure 1

THE REENTRY TRAJECTORY

Chapter One pointed out that there is no "top" to the atmosphere of the earth. The atmosphere is most dense at the surface of the earth, and density decreases as altitude increases. But no matter how far the spacecraft travels from the earth, density never becomes "zero." Consequently, spacecraft hundreds of miles above the earth's surface develop an insignificant amount of aerodynamic drag. Reentry into the atmosphere begins at the altitude where drag first becomes significant. This altitude depends primarily upon vehicle velocity and physical characteristics. For a Mercury spacecraft, reentry begins at about 250,000 ft above the surface. For an Apollo spacecraft returning from a lunar mission at 25,000 mph, reentry begins at 400,000 ft.

The spacecraft trajectory during reentry results from the effects of gravitational attraction of the earth and aerodynamic drag; these forces are affected by:

1. Entry velocity, which depends on the spacecraft mission.
2. Atmospheric density, which is a function of altitude.
3. Entry angle, the angle with the local horizontal at which the vehicle enters the atmosphere.
4. Spacecraft characteristics:
   a. Weight (or mass).
   b. Shape of the vehicle (how streamlined it is).
   c. Area of the vehicle, measured at its maximum cross-section.
   d. Lifting characteristics. Some spacecraft are ballistic and develop no lift force in the atmosphere (ICBM warheads and the Mercury capsule). Other spacecraft develop lift on reentry (Gemini, Apollo, and lifting body spacecraft).

The forces acting on an unpowered spacecraft are shown in Figure 2. The weight force depends upon the mass of the spacecraft and its altitude. The force acts toward the center of the attracting body (the earth), tends to accelerate the spacecraft, and curves its trajectory toward the vertical (entry angle of 90°).

Drag force results from aerodynamic resistance to motion through the atmosphere. The magnitude of the force depends upon atmospheric density, vehicle velocity, and vehicle characteristics. For example, a streamlined vehicle like an ICBM warhead would develop less drag on reentry than would a blunt spacecraft like Mercury, Gemini, or Apollo, if other factors are the same.

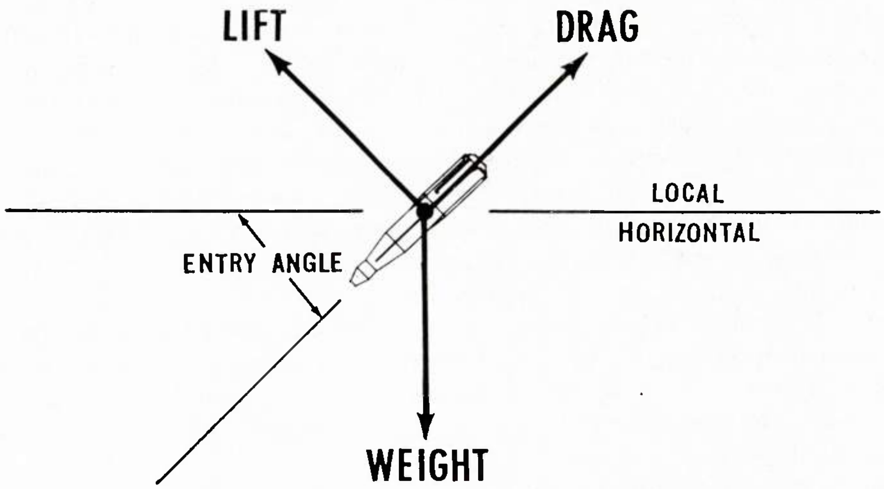


Figure 2

Similarly, the lift force depends upon atmospheric density, vehicle velocity, and vehicle characteristics. The spacecraft must be designed to develop lift during reentry if a lifting entry is desired. Otherwise, it will make a ballistic reentry.

A ballistic reentry or trajectory is, by definition, one in which the only forces acting on the spacecraft are weight and drag; it is unpowered (no thrust force) and no lift is developed. In a ballistic trajectory, the path is fixed, and the vehicle cannot be maneuvered from this fixed path. Thus, it cannot make corrections to insure hitting the preplanned touchdown point.

In a lifting spacecraft, however, the lift force can be varied during reentry to allow the spacecraft to maneuver (change its trajectory) and make path corrections toward the preplanned recovery area.

In summary, the reentry phase of the spacecraft trajectory serves a useful purpose. During its passage through the atmosphere, the spacecraft generates aerodynamic forces. The drag force allows the vehicle to be slowed to a soft landing on the earth's surface without the use of a retrorocket. In addition, if the spacecraft is designed so that lift is developed, it can maneuver during reentry and change its trajectory and landing point.

These advantages are not obtained free, however. There are also problems associated with reentry.

Problems Caused During Reentry

Two major problems occur during the atmospheric reentry of a spacecraft:

1. Heating of the spacecraft.
2. Deceleration loads on the spacecraft and payload.

The heating problem arises because the reentering spacecraft possesses a tremendous amount of kinetic energy. Kinetic energy is the ability to do work by virtue of the velocity the spacecraft has. For example, the Apollo spacecraft returning from a lunar mission weighs about six tons and travels at approximately

36,000 feet per second or 25,000 miles per hour, giving it about 486 billion foot-pounds of kinetic energy. As the spacecraft is slowed by the aerodynamic drag of the atmosphere, the kinetic energy is transformed into thermal energy in the atmosphere surrounding the spacecraft. This energy transfer takes place through impact of air molecules with the spacecraft and its shock wave.

Because the spacecraft is surrounded by an envelope of heated air, heat is transferred to the spacecraft structure by radiation and convection processes. The amount of heat transferred to the spacecraft depends upon many different factors: entry velocity, atmospheric density, characteristics of air flow around the spacecraft, reflectivity of the spacecraft surface, and others.

Since this transfer process is fairly complicated, an accurate prediction of the maximum temperature that can be expected during a particular reentry is very difficult. However, if all the heat in the atmosphere (or even a large percentage of it) were transferred into the spacecraft, kinetic energy transformed to thermal energy would vaporize the spacecraft like a meteor.

Two heating values must be calculated for each reentry—the maximum heating rate and the maximum total heat load. The maximum heating rate normally occurs at the leading edge or stagnation point of the spacecraft, and the peak temperature reached during reentry occurs at this point. Therefore, the nose of the spacecraft usually has the most critical heat protection requirements. Heating rate is expressed in units of *B*ritish *T*hermal *U*nits (BTU's) per square foot per second and is a measure of how rapidly the heat is being transferred from the atmosphere into some representative area of the structure.

Since each type of material has a maximum temperature at which it can sustain the structural loads placed on it, the type of material used for different parts of the spacecraft structure must be selected carefully according to the expected maximum heating rate at that part during reentry. The maximum heating rate will depend upon how rapidly kinetic energy is changed to thermal energy on reentry and also on the type of heat protection used to decrease heat transfer from the atmosphere to the spacecraft structure.

If one spacecraft is heavier or reenters the atmosphere at a greater velocity than another, it will have greater kinetic energy and, if other conditions are the same, the maximum heating rate at any point on the structure will be greater. For example, Apollo is heavier and reenters at a greater velocity from its lunar mission than Mercury did from its earth orbital mission. Therefore, Apollo has greater kinetic energy, a more severe reentry heating problem, and a heavier heat protection system than Mercury.

The other heating problem, the total heat load, is measured in BTU's per square foot. This is the total amount of heat which reaches any particular area on the spacecraft during the entire reentry. Again, the heat protection system must withstand this load to be effective. Heat load depends upon kinetic energy lost during reentry and heat transfer characteristics between the spacecraft and the atmosphere.

The trajectory will also affect reentry heating. If a spacecraft reenters at a relatively steep angle (an ICBM warhead, for example, reenters at approximately 20°), the heating rate and stagnation temperatures tend to be high. However, since the vehicle passes quickly through the reentry phase without time to absorb

much of the thermal energy generated, the total heat load tends to be relatively low.

On the other hand, a manned ballistic spacecraft like Mercury reenters the atmosphere at a relatively shallow angle around 1-2°. Therefore, reentry heating is spread over a longer period of time. This causes the maximum heating rate to be relatively low, but since the vehicle is in the heating environment for a longer period of time, the total absorbed heat load tends to be higher.

The deceleration problem arises because the spacecraft is being slowed from the reentry velocity by aerodynamic drag. Deceleration, which is a negative acceleration or a slowing down, is the time rate of change of velocity. The average deceleration during some interval is the ratio of the velocity change to the time interval within which it occurs. Deceleration is usually measured in g's, the ratio of the deceleration to the acceleration of gravity at the surface of the earth—32.2 feet per second per second. An astronaut undergoing a deceleration of 10 g's would "weigh" ten times what he weighs on the surface of the earth.

Both astronauts and spacecraft have limits of deceleration which they can survive. Equipment can be designed to withstand very high decelerations (up to hundreds or even thousands of g's in some cases), but such designs usually mean an increase in weight. Man, however, cannot be redesigned and must avoid a reentry involving high decelerations.

The deceleration limit used for manned spacecraft reentry is usually 10 g's applied from astronaut chest to back because this is the maximum deceleration at which a crewmember is effective. Man can survive reentry up to 20 g's deceleration, but this limit is applied to emergency situations only. Because of this, the Apollo spacecraft is also designed to survive a 20 g deceleration reentry.

Deceleration on reentry depends upon vehicle velocity, atmospheric density, vehicle characteristics, and entry angle. The greater the vehicle velocity, the greater the aerodynamic drag slowing the spacecraft, and other things being equal, the greater the deceleration. Similarly, a blunt-shaped spacecraft also develops greater drag than a streamlined spacecraft. Finally, if a vehicle such as the ICBM warhead enters at a steep angle, it will descend at a high velocity into the lower denser atmosphere where aerodynamic drag is greater. Its deceleration is greater than that experienced by a Mercury spacecraft which entered at a shallow angle and slowed down high in the less dense part of the atmosphere. Therefore, a warhead experiences decelerations of about 150 g's whereas Mercury reentered with a peak deceleration of approximately 8 g's.

A third reentry problem arises if there is a requirement for maneuvering the spacecraft during reentry to arrive at a precise touchdown point on the surface of the earth, to obtain a large entry corridor, or to obtain a large area for possible landings.

If the spacecraft develops no lift force on reentry, then it follows a ballistic trajectory and no control over this trajectory during reentry is possible unless rocket thrust is used. Timing and position of retrofire used to bring it out of orbit was the only control that Mercury had over its touchdown point. At the Mercury orbital altitude of 100 NM, the aerodynamic drag was sufficient to cause the orbit to decay into the lower atmosphere and bring the spacecraft back to earth in

about 7 days. However, Mercury carried three small retrorockets which, when fired, decreased velocity by approximately 500 feet per second so that the spacecraft would return to earth in less than one orbit and have some control over the location of its landing point. Once retrorocket firing was completed, the Mercury astronauts had no further control over the trajectory. Since orbital velocity at retrofire was 24,000 feet per second, an error of one second in timing meant an error of 4 NM at touchdown; timing was critical.

If an ICBM warhead does not reach orbital velocity, it does not require a retrorocket to bring it back through the atmosphere to the surface of the earth. However, a means of maneuvering during reentry allows the warhead to correct its trajectory and to avoid enemy interception.

For future spacecraft, maneuvering during reentry will allow more flexibility in spacecraft operations. A lifting body spacecraft will be able to deorbit from space to a particular runway from many different points in its orbit around the earth. Recovery forces sent out to locate and retrieve spacecraft can be decreased in size and cost.

From the foregoing analysis, major problems obviously associated with reentry have been overcome in past space operations and must be anticipated in future operations.

Solutions to the Problems of Reentry

The magnitude of the heating problem on reentry not only may limit the range of reentry trajectories available but also requires that spacecraft be equipped with some type of heat protection. The preceding section pointed out the relationship between heating and entry angle. A shallow cntry angle for a given mission will result in lower heating rates within the atmosphere and at the spacecraft surface and in higher total heat loads on the spacecraft than those of a steep reentry. A heat protection system decreases heat transfer between atmosphere and spacecraft or transfers heat back to the atmosphere instead of allowing it to enter the vehicle structure.

The shape of the reentry vehicle affects heat transfer. Vehicle shape determines the standoff distance of the shock wave (distance between shock wave and leading edge) created during entry into the atmosphere. The greater the standoff distance, the less the heat transfer across the shock wave to vehicle surface.

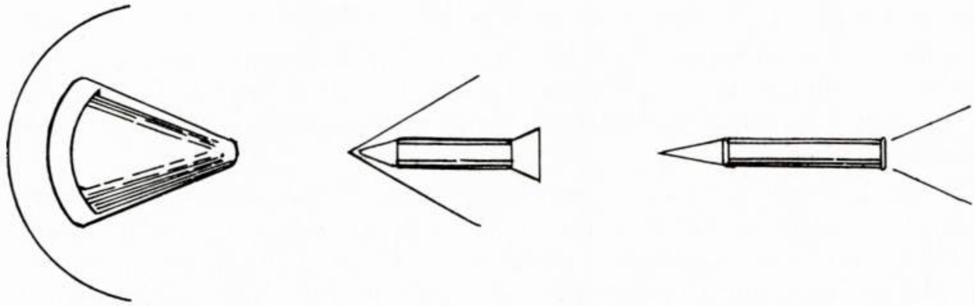


Figure 3

The air between the shock wave and the blunt-shaped spacecraft is moving more slowly than the free flow. The boundary layer, a thin layer of air next to the vehicle, moves very slowly. This slowly moving air serves as an insulating layer to decrease heat transfer. On the streamlined vehicle, the shock wave is closer to the vehicle; the air is moving faster; and heat transfer is greater. On the needle-shaped vehicle, the shock wave is located at the aft end and affords little or no heat protection.

Many different techniques have been used, and are presently being developed for future use, to provide further heat protection during reentry. Most of these can be classified into six different types: heat sink, ablation, insulation, radiation, transpiration, and spray cooling. The general characteristics of these systems are as follows:

1. The *heat sink* provides heat protection through the addition of a large mass of material (copper in early missile warheads) which absorbs a great quantity of heat as it warms up. There must be enough mass to handle all of the heat of reentry prior to the material melting or boiling away. Unless the material has some other function after reentry, such as providing fuel, then it becomes a payload penalty. Today, the heat sink is used for component heat protection but not for cooling entire reentry vehicles.

2. The *ablation* heat protection system has been used on all U. S. manned spacecraft and on most of the warheads in the inventory. Ablation is the wearing away, melting, charring, and vaporizing of the surface material. Most ablation materials are of the melting or charring type. The melting types are generally undesirable for reentry heat protection because the melted material tends to run and change the aerodynamic shape of the vehicle. This could have serious effects at high velocities of reentry. Charring ablation materials have been used on most reentry vehicles. These materials are complex organic mixtures; reinforced plastics, epoxy resins, fiber glass, and others. Ablation systems are very efficient heat protection methods comprising approximately 10% of the reentry vehicle weight. Reentry heat is handled by five different methods. As the ablation material is heated during reentry, it undergoes a chemical reaction which results in charring and formation of a gas:

a. Heat is absorbed in the chemical reaction;

b. The gas formed "outgasses" (flows out) from the surface, becomes heated, and carries this heat downstream;

c. The gas, flowing along the surface, provides boundary layer insulation against convective heating;

d. The ablative material is a poor conductor of heat (a good insulator) so it stores heat by getting hot without readily conducting it into the interior of the vehicle; and

e. The surface becomes red hot which causes heat to be re-radiated back to the atmosphere.

Because of these five methods, ablation is efficient for heat protection. Its main disadvantage lies in the difficulty of using ablative shielding for more than one reentry without extensive refurbishment.

3. The *insulation* technique involves the use of a material that is a poor heat conductor. Although this material heats up on reentry, it does not conduct much of the heat into the interior structure. Because of this, the material will become very hot and, therefore, must be capable of surviving high temperatures. Ceramics have been used for insulating the nose cap on unmanned test vehicles such as Asset. The insulation method can be used with, and is related to, the radiation method.

4. *Radiation* of heat from the spacecraft to the atmosphere can be used to dissipate the heat of reentry. With this method, the surface of the spacecraft is allowed to heat up on reentry until it gets so hot that the equilibrium temperature is reached. At this temperature, heat coming in by convective and radiative heat transfer equals the heat being re-radiated to the atmosphere. The material used must withstand the equilibrium temperature. For a steep ICBM warhead type of reentry, there is no known material that can survive for long the resulting high heating rates and temperatures. For reentry of a lifting body spacecraft which has relatively low heating rates and low equilibrium temperatures and spends a long time in reentry for reradiation of the heat, the radiation method appears economical. Even here, the temperatures become sufficiently high to require the use of rather exotic structural materials, such as graphite, columbium, and molybdenum which are difficult to form and may present oxidation problems at high temperatures. Since these materials are not consumed during reentry, they appear attractive for future multiple-reuse reentry spacecraft. Coating materials to prevent oxidation must first be developed.

5. *Transpiration* heat protection requires that a liquid be boiled off at the surface of the spacecraft so that heat is absorbed in the change of state and is carried away in the vapor. This limits surface temperature to the boiling point of the liquid just as water in a teakettle keeps the teakettle at the boiling point of the water, even over a flame. The major disadvantage of this method is that it is an active system and extensive plumbing would probably be required to get the cooling material to the outer surface. A failure in this system during reentry would be disastrous. Transpiration cooling is somewhat related to spray cooling. Both would probably use water or lithium as the cooling substance.

6. *Spray cooling* also utilizes the boiling off of a liquid, but, in this case, the liquid is sprayed on the surface to be cooled in liquid droplets that vaporize on contact. If this is done on the inner surface, the vapor could be collected and sent through a heat exchanger located at a cooler part of the vehicle. This system could be used on the nose cap or leading edge of a reusable reentry vehicle, and the coolant reservoir would merely have to be refilled to refurbish the heat protection system.

The deceleration problem on reentry can be controlled only by controlling the rate of velocity change as the spacecraft is slowed by aerodynamic drag. In the case of an ICBM warhead, slowing down in the atmosphere makes it more vulnerable to enemy action; its shape and trajectory are designed so that the warhead is slowed as little as possible on reentry. If warheads are properly designed, deceleration is not an operational restriction imposed on the trajectory.

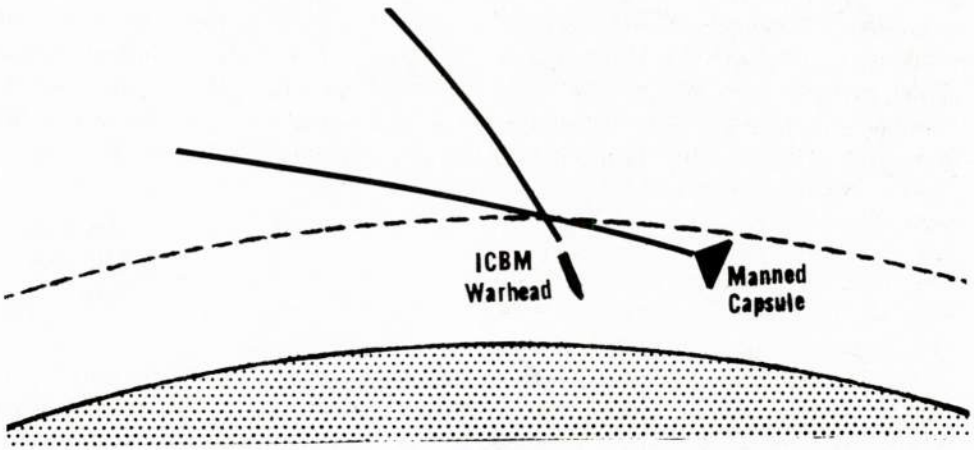


Figure 4

In the case of a manned spacecraft, reentry deceleration is limited to 10 g's. Since the astronaut is being slowed at the same rate as the capsule, there is no way to insulate the astronaut from deceleration as is done with heating. Because a soft landing is required, which means essentially a "zero velocity" at touchdown, the total velocity change during reentry is approximately equal to the entry velocity. Therefore, the only way to control deceleration is to control the rate of velocity change.

If a shallow entry angle is used, the spacecraft begins slowing down higher in the atmosphere and takes longer to reenter, with a resulting lower peak deceleration. This is why a shallow entry angle is used for returning manned capsules to earth, as shown in Fig. 4.

To overcome the problems of heating and deceleration, spacecraft have been reentered into the atmosphere at a shallow angle to spread the effects of heating and deceleration (kinetic energy change and velocity change) over a longer period of time. If the spacecraft develops a lift force in the atmosphere, it can use this lift to control the rate of descent and also to maneuver during reentry, helping to solve the third problem. For these reasons, a great deal of research is presently being done on lifting reentry spacecraft.

Maneuvering on reentry can be best analyzed by considering two different portions of the reentry trajectory—the high velocity portion from reentry altitude and orbital velocity down to 100,000 feet altitude and Mach 2 (twice the speed of sound) and the low velocity portion from 100,000 feet to touchdown on the surface.

In the high velocity part of the trajectory, there are essentially two ways of maneuvering the spacecraft—by using lift or by using rocket engines. Rocket engines are fairly heavy devices and have not been used for reentry maneuvering of manned spacecraft. Therefore, lift is the primary means of providing a maneuvering capability during reentry.

Lift, like drag, is an aerodynamic force generated within the atmosphere. A lifting capability is of no use in space but is merely a weight penalty because of the additional structure required. Once the spacecraft encounters sig-

nificant atmospheric density on its descent, a useable amount of lift can be generated. Although Mercury was a ballistic spacecraft which developed no lift on reentry, both the Gemini and Apollo spacecraft (with symmetrical shapes similar to Mercury) do develop a small amount of lift. Lift is generated by building and loading these vehicles so that the center of mass (or center of gravity) is offset from the horizontal centerline as shown below:

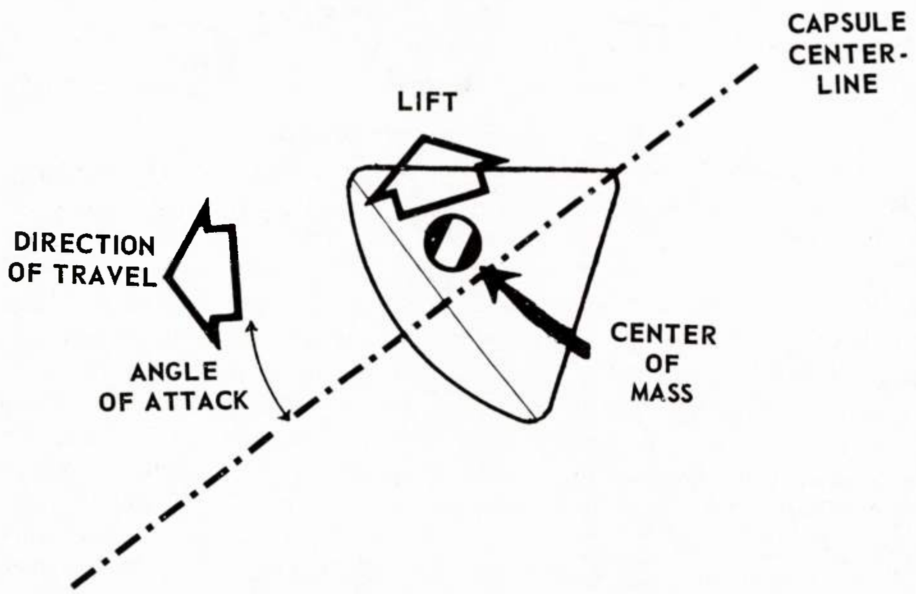


Figure 5

This causes the spacecraft to be aerodynamically stable during reentry at an angle to the direction of travel (angle of attack). This, in turn, generates a pressure differential and a lift force, just as an aircraft wing generates lift. By using the attitude control reaction engines to rotate the spacecraft, the astronauts can point the lift vector in the direction they want to travel to correct the trajectory. A lifting body spacecraft generates lift by virtue of its unsymmetrical shape.

In the low velocity portion of the reentry trajectory from the 100,000 ft altitude and about Mach 2 down to touchdown, many different means of generating lift for maneuvering to a soft landing have been tested. On early and present manncd spacecraft, combinations of small drogue parachutes and large main chutes were used to lower the capsule gently to the water with no horizontal maneuvering possible. For thc future, such devices as lifting parachutes, flexible paraglider wings, deployable rotors, and lifting body spacecraft with sufficient subsonic lift to make runway landings are being tested. These devices will allow horizontal maneuvering to a preselected landing point and a land recovery to be made. The large sea recovery forces presently used may not have to be deployed in the future.

In summary, the advantages of having lifting capability in the reentry spacecraft are:

1. Low heating rates;
2. Low deceleration loads; and
3. Ability to maneuver in the atmosphere

Disadvantages are:

1. Lifting surfaces add weight which reduces useful payload in space;
2. The total heat load on reentry is greater because the vehicle spends more time in the heating environment on reentry; and
3. A lifting entry is more complex than the simple ballistic trajectory with its one possible path.

In this chapter, we have considered some of the problems that arise during reentry and some of the solutions that have been used or are under study for future use. Obviously, reentry is a complex subject. Only a simplified, basic treatment of the subject has been given here. A great many, more rigorous and complete treatments of the subject are available. These problems and their solutions are presently under continuing intensive study by space engineers and scientists.

BIBLIOGRAPHY

DUNCAN, ROBERT C. *Dynamics of Atmospheric Entry*. New York: McGraw-Hill, 1962.

ELY, LAWRENCE D. *Return from Space*. Springfield, Ill.: Charles C. Thomas, 1966.

GAZLEY, CARL, Jr. "Decleration and Heating of a Body Entering a Planetary Atmosphere from Space" in Air Force. Office of Scientific Research. *Astronautics Symposium*. 1st- 1957- . New York: Pergamon Press, Symposium Publications Division, 1958. (Volume title: *Vistas in Astronautics)*

LA GALLEY, DONALD P. *Ballistic Missile and Space Technology Re-entry and Vehicle Design, Vol. IV*. New York: Academic Press, 1960.

LOH, W.T.H. *Dynamics and Thermodynamics of Planetary Entry*. Englewood Cliffs, N.J.: Prentice-Hall, 1963.

MORROW, C. T., ELY, L. D., and SMITH, M. R. *Ballistic Missile and Aerospace Technology Re-Entry*. New York: Academic Press, 1961.

RAND Corporation. *Aerodynamic Aspects of Future Space Vehicles and Implications for Testing*. Washington: RAND Corporation, 1967.

———. *Atmospheric Entry of Manned Vehicles*. Santa Monica, Cal.: RAND Corporation, 1960.

———. *An Introduction to Astronautics, Vol. I*. Santa Monica, Cal.: RAND Corporation, 1958. Pp. 10.1 through 10.32.

TRUITT, ROBERT W. *Fundamentals of Aerodynamic Heating*. New York: Ronald Press, 1960.

Appendix H

COMPUTERS IN SPACE DEFENSE

THERE is one military space operation—the Space Defense System—that would be nonexistent if it were not for the capability of a computer system. This operation will be reviewed as an example of an application of computers to our military space capability. Moreover, this system's operation will serve to point out the myriad to tasks that may be accomplished by digital computers.

Computers in Space Defense

The Space Defense System is comprised of the total capabilities and facilities provided to CINCNORAD/CINCONAD to detect, track, and identify all man-made objects in space; to provide warning of potential or actual space threats; and to provide a defense against space systems. The command, control, and computational center of the Space Defense System is known as the Space Defense Center (SDC). The SDC is the hub of a worldwide network of radar, radio, and optical sensors which maintain surveillance on all man-made objects in space, and it integrates all elements of the Space Defense System. The technical operation of the SDC computer and communications facilities is the responsibility of an Air Force organization—the 1st Aerospace Control Squadron of the Aerospace Defense Command.

Since 1957, close to 4,000 man-made satellites have been placed in orbit. The computer capability of the SDC has also increased to keep pace with the satellite population growth and increased accuracy requirements. The SDC computer system has evolved from a 10-microsecond memory access, nonreal-time input and output system in 1961 to a 1.5-microsecond memory access and a real-time input/output with interrupt features. Today's operational computer system uses the Philco S-2000, 212 model computer and is known as the Delta 1 system.

The Philco S-2000 central processor is a high-speed, large scale asynchronous, electronic digital computer which consists of a control section that interprets and executes program instructions plus 32,000 (32K) words of magnetic core storage.

The control section is housed in one hardware module. It consists of an instruction unit, an index unit, an arithmetic unit, and a store unit. The 32K-core storage is a random access device found in four hardware modules each containing 8K words of magnetic core memory. Each word consists of 48 bits plus eight odd parity bits. A word may contain data or instructions. The magnetic core memory has an average access time of 0.9 microseconds and a complete read-write cycle time of 1.5 microseconds. The control section normally executes program instructions (stored in core memory) in sequential order.

The instruction unit accesses up to four instructions from memory until they can be accepted by the index unit. It contains two program address registers and two 48-bit program registers.

The index unit obtains operands and stores them until the arithmetic unit has completed the preceding instruction. There are eight index registers, each consisting of as many bits as are needed to address the largest memory address.

The arithmetic unit receives operands and instructions from the index unit, executes the instructions, and transfers the results to the store unit.

As part of the Central Data Processor, there is also a Real-Time System, magnetic tape and magnetic drum storage, and an Input-Output Data Controller (IODC) with associated equipment. Although the input/output and communication interfaces are significant computer operations in themselves, this review will be limited to the Delta 1 program operation and data flow.

Oversimplified, the data flow in the SDC begins when a sensor site observes a satellite and transmits the observation to the SDC. When the observation is received, the system compares that observation with the orbital elements of all cataloged satellites in an attempt to identify the satellite. The orbital elements are stored in the computer. If the observation is identified as a known satellite, certain routine actions will occur. If the observation is not on an identified satellite, then a set of orbital elements must be generated. These orbital elements become the basis for the majority of outputs required for operation of the system. They are used to produce look angles—the pointing data which enable sensor sites to acquire an object in order to obtain more observations. Many sensors have on-site computer facilities. Those sites would receive orbital elements from the SDC and, using their own computer, produce look angles.

Bulletins, the ephemeris or predicted position information for a particular satellite, are another major output which are provided to many agencies in addition to Space Defense System organizations.

The entire flow of data operates in a closed loop. Observations from the sensors flow into the SDC. In turn, the SDC provides orbital elements or look angles to the sensors as a reacquisition aid. When more observations are received from the sensors, the SDC computer system differentially corrects the data. This correction process is a method for finding small corrections from the differences between the observed and computed coordinates (residuals). When applied to the elements, the small corrections reduce the deviations from the observed motion to a minimum. This cycle iterates continuously until the satellite decays or is deorbited. Selected computer processes and sequences that enable this flow to occur are the subject of the remainder of the section.

The Delta 1 program operates in response to manually or automatically entered stimuli (Fig. 1). The operation of the system is controlled by a program executive which provides for: (1) accepting, monitoring, and relaying real-time observation message inputs; (2) selecting certain processes automatically; (3) accepting operator/analyst requests for process operation and data insertion; and (4) operating processes according to priority.

The program is structured into three levels of responsibility (Fig. 2): *level one* consists of functions which control the routing of input/output data, schedule the activities of the program, and provide the interface mechanism by which the operator communicates with the program; *level two* is the process or job-related control which directs the execution of particular processes or jobs to be run; *level three* consists of the individual functions.

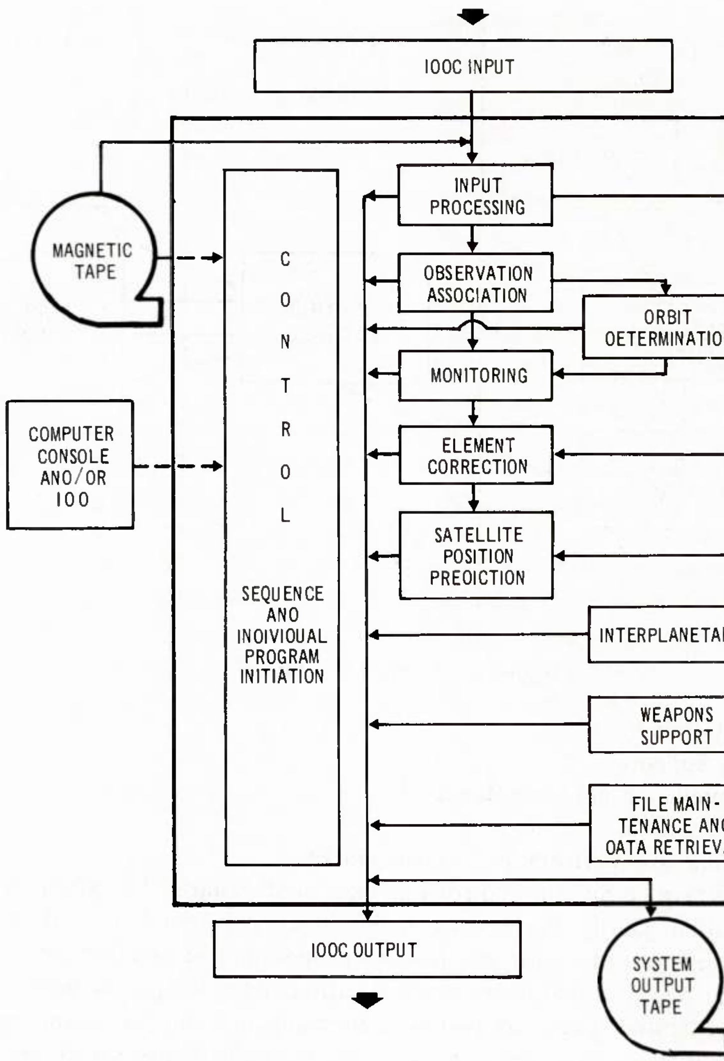


Figure 1

System Processes

A process may be a single programmed function or a prescribed sequence of functions operated by the executive according to the stimulus and the type of space object concerned. In the Data Processing Subsystem, some of the Delta 1 System functions are grouped in the following areas:

a. System Control
b. Input Processing
c. Observation Association
d. Observation and Element Monitoring
e. Element Correction
f. Satellite Position Prediction

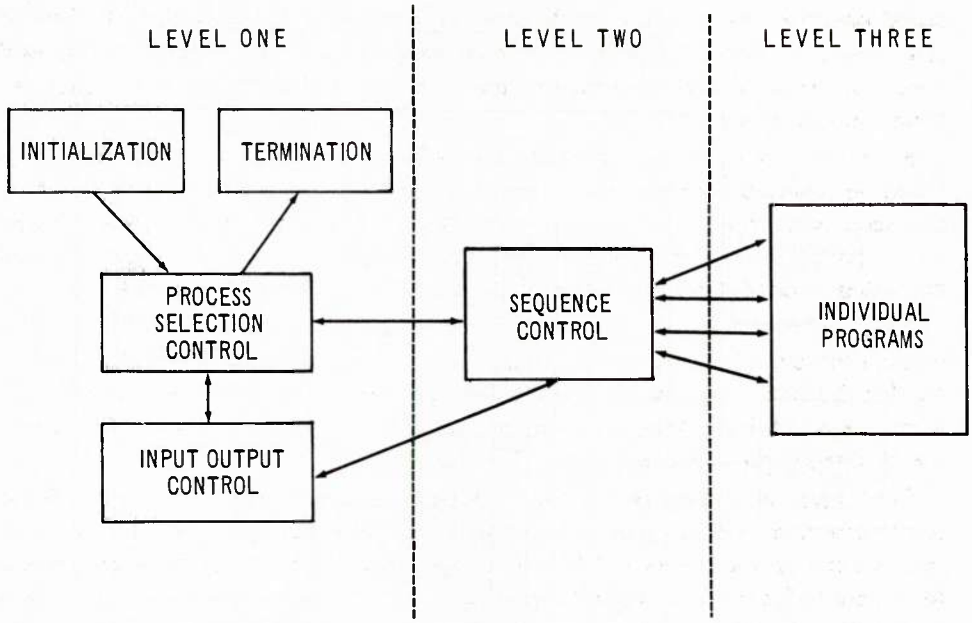


Figure 2. Control structure.

g. Orbit Determination
h. Weapons Support
i. File Maintenance and Data Retrieval

Briefly, these functions perform the following tasks.

SYSTEM CONTROL.—System control initializes and restarts the system to allow program operation. During the operation, this function responds to certain equipment malfunctions in order to insure continued operation, if possible, on degraded equipment. It also insures that operations are terminated at the proper time.

INPUT PROCESSING.—There are two basic functions of input processing: processing satellite observation messages and processing operator inputs. Observation message processing includes legality checking of the observations and converting them to a standard format. These observations are then retained for use in the observation association function covered below. Operator input processing allows manual data input via a schedule tape,* the input output data controller (IODC) and/or the computer console.

OBSERVATION ASSOCIATION.—The observation association function correlates the observations received from input processing with cataloged element sets. There is a predetermined association criteria which establishes the limits for correlation. According to these limits, observations are classified into one of the following categories: associated, doubtful, or unassociated.

MONITORING.—This function monitors observations and element sets. After association processing, the monitoring function sorts and merges newly categorized observations into intermediate storage piles for later use. Associated or doubtfully asso-

* Schedule tape—a magnetic tape which contains prestored input parameters for one or more programs and/or sequences to be run in succession.

ciated observations are used for element set monitoring. Unassociated observations can be used in the orbit determination function covered later. Element sets are monitored for excess divergence if the satellite is suspected of being lost or decayed, or if a new bulletin is required.

ELEMENT CORRECTION.—Element correction improves an object's element set based on reported observations of that object. The function operates as part of several sequences or as an individually requested function. As an individual function, any of several general or high-precision orbit correction techniques may be manually requested to correct selected input element sets with selected input observations.

As a part of a sequence, element correction, using a weighted least-squares technique, differentially corrects the element sets which are designated manually or by another function (e.g., monitoring), using the observations in the system files. Element sets which have been corrected as part of a sequence are retained for possible use by satellite position prediction.

SATELLITE POSITION PREDICTION.—Satellite position prediction uses element sets to generate predictions in various forms, quantities, and groupings for Space Defense Center requirements and for other users of this data. Predictions are prepared for output to local personnel or for transmission. The look-angle surveillance predictions used by sensors are, in general, produced automatically following element correction or the determination of a new element set. The other prediction functions are usually operated by request on manually designated or input element sets.

The following types of predictions are generated:

a. Sensor Acquisition Data—Satellite acquisition coordinates or element sets for use by specific sensors for observing and reporting on known satellites.

b. Satellite Bulletins—Data representing the nodal crossings, ground trace, and altitude of the satellite over a period of time.

c. System Function Requirements—Satellite positions(s) for use by a program system function.

d. Ephemeris Generation—Satellite position predictions at set time intervals over a period of time; varied formats and data quantities accommodate several users of this data.

e. Position Reports—Position situation reports showing the status of satellites at a given time, within a given time interval, or within a given volume of space.

f. Orbital Intersection—The closest point of approach of two satellites.

ORBITAL DETERMINATION.—Orbital determination produces an initial set of orbital elements for the system catalog. These elements are produced in one of three ways: (1) computation of an orbit directly from the observations of an unknown object reported by a sensor; as many as four techniques may be used concurrently, depending upon the types of observations received; (2) correlation of an unknown track (input observation set) from certain sensors with historical sensor data on previously launched objects; and (3) computation of an orbit directly from prelaunch data of a flight path. For the first two methods, orbit determination operates automatically when the association of function confirms there is a tagged unknown observation track; the third method is manually initiated. Following orbit determina-

tion, the association, element correction, and prediction functions are operated to improve the initial orbit and to provide information on the new orbit path.

WEAPONS SUPPORT.—The weapons support function provides satellite intercept planning and prediction information.

FILE MAINTENANCE AND DATA RETRIEVAL.—This function permits the operator to access and change the system data files. Each maintenance and retrieval function is manually requested.

Sequence Control

A sequence controller in the Delta 1 system is a Level 2 executive program which operates individual programs in a prescribed order. Generally, the sequence controllers prepare the data environment for an individual program operation, establish input/output options, load individual programs which will be operated, and transfer control to that program as well as determine when conditional programs will operate. Following are some of the operations of the sequence controllers (sequence is normally written as SEQ followed by the number of the sequence, e.g., SEQ01 = Sequence 1):

Sequence 1—Observation Input Processing.

This sequence provides for the basic conversion, association, and maintenance of sensor observation reports. System control accumulates these observation reports in the observation input buffer on drum, and on a tape buffer when the drum has filled. The operation of SEQ01 clears all requests for SEQ01 from a process request queue.

Sequence 2—Routine Cataloging

This sequence filters the observation files to optimize their currency and usefulness. It also monitors element sets to select those requiring differential correction, obtains corrected elements, and generates new acquisition parameters.

Sequence 3—Tagged Unknown Processing

This sequence provides automatic processing of tagged unknown observations. System control stores these observations in a special observations input buffer and requests the operation of the sequence either when the last observation in a set is received, or ten minutes has elapsed since the receipt of the first observation in a set.

Sequence 4—Observation Insertion

This sequence will provide for the insertion of standard observation cards into the system files and for the operation of the report association program* in a retro-association mode, with the file of unknown observations. The sequence can be used either to insert analyst-prepared observations into the system with a predetermined

* Report Association Program—A program that attempts to associate all incoming observations with element sets.

satellite number, or observation reports which had been corrected/converted elsewhere.

Sequence 5—Satellite Renumbering

This sequence provides the capability to replace a temporary satellite number with a permanent satellite number in the system files. The observations for a specified satellite are retrieved from intermediate and master observation files and placed in a drum storage area. The satellite number is changed on the element set, and the element file is written onto the data files. Thc satellite number is changed on the observations from the drum storage file and written onto the intermediate observation file. The observations with the temporary satellite number are then purged from the master observation file.

Sequence 6—Analyst Report Association

This sequence provides for insertion of an unlimited amount of taped observations to an individual program—the Analyst Report Association program.

Sequence 7—Tagged Observation Insertion

This sequence provides for tagged observation input on punched cards and for merging them onto the intermediate observation file. Residuals are computed and inserted by an observation residual calculation program.

Sequence 10 through 15

There is a series of sequences which provide a capability to produce a number of combinations of basic outputs required for the SDC operation. These sequences perform their operation on satellites that are manually designated by SDC personnel. SEQ10 differentially corrects (DC) those satellites and updates the element file when a correction is successful. SEQ11 provides a DC, an element update, and generates a set of look angles. SEQ12 includes all SEQ11 outputs plus a bulletin. SEQ13 will provide a DC an element update and a bulletin. SEQ14 will generate and provide for transmission of look angles. SEQ15 adds a bulletin to the SEQ14 output.

Sequence 17—Observation Editing and Differential Correction

This sequence provides for differentially correcting orbital elements of manually designated satellites using observations from the intermediate observations file and the master observation file.

Sequence 20—Sensor Tasking

This sequence provides the capability to update tasking codes, transmit tasking messages, and generate tasking summaries. This information establishes priorities and requirements for sensor sites.

Sequence 21—Domestic Launch

This sequence facilitates the cataloging of U.S. launched satellites by establishing a tentative element set and providing ephemeris and acquisition data to sensors prior to the actual launch. Prelaunch parameters and recipient sensors are designated by card input.

Sequence 22—Manual SEQ03

This sequence provides for determining an initial orbit from manually input observations, differentially correcting the element set, updating the system files, and alerting the sensors. This sequence is essentially the same as Sequence 3 (Tagged Unknown Processing).

There is a wide variety of individual programs which are part of these sequences (as well as a vast number not included in these sequences) which produce the results needed for a space defense operation. It should be clear that man could not accomplish the mission without the aid of a high-speed digital computer.

INDEX

This is designed to be a working index. It differs from the normal alphabetical index in that the most appropriate word (rather than the first) in an item has been catalogued. In most cases, the noun has been chosen for listing with the modifying adjective shown as a subheading. For example, Reentry Angle is listed as:

ANGLE
    Reentry

and Mass Ratio is:

RATIO
    Mass

GLOSSARY OF SYMBOLS

1. a semi-major axis of ellipse; average linear acceleration
2. b semi-minor axis of ellipse
3. bo subscript for burnout conditions
4. c distance between focus and center of ellipse
5. e base of natural logarithm, = 2.718
6. g local acceleration due to gravity, = 32.2 ft/sec^2 at the surface of the earth
7. h altitude, height above surface of earth
8. h_a altitude of apogee
9. h_p altitude of perigee
10. i angle of inclination
11. p electric power
12. r radius length; mixture ratio
13. r_a radius to apogee
14. r_e radius of earth, = 3440 NM = 20.9×10^6 ft
15. r_p radius to perigee
16. s linear displacement
17. t time in sec
18. u eccentric anomaly
19. v linear speed, velocity magnitude
20. v_e nozzle exit velocity
21. Δv increment or change of speed
22. w work in ft-lb force
23. A area
24. A_e nozzle exit area
25. C_D coefficient of drag
26. C_L coefficient of lift
27. D atmospheric drag
28. E specific mechanical energy in ft^2/sec^2

29. F focus of ellipse; force in lbs; thrust in lbs

30. G Universal Gravitational Constant, = 10.69×10^{-10} ft^3/lb mass-sec^2

31. H specific angular momentum in ft^2/sec

32. I_{sp} specific impulse in sec

33. I_t total impulse in lb-secs

34. KE kinetic energy

35. L aerodynamic lift; latitude

36. M mass in slugs

37. $\dot{M}$ ıass flow rate

38. NM nautical miles, = 6080 ft

39. P period of revolution; pressure

40. PE potential energy

41. Q ballistic trajectory parameter; reactor thermal power

42. W weight in pounds force

43. $\dot{W}$ weight flow rate

44. α average angular acceleration

45. δ very small change or error

46. ε eccentricity; expansion ratio

47. η electrical efficiency

48. θ angular displacement

49. λ longitude

50. μ gravitational parameter, = 14.08×10^{15} ft^3/sec^2 for earth

51. ν true anomaly; heat transfer efficiency

52. π conversion constant, = 3.1416; π radians = 180°

53. ρ atmospheric density, $slugs/ft^3$

54. ϕ flight path angle, elevation angle of velocity vector

55. ψ free flight range angle

56. ω argument of perigee; average angular speed

57. Δ increment of

58. Ψ thrust to weight ratio

AU-GAFB. ALA (703217) CATO SHOW PRINT, CATO, N.Y., 10/70, 4500